

[illegible]

67 easnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnT 84  
 7912 CAATAAATATGGAACCCAGCAACCAATATCGGATGTTTCCAAACT 7863  
 84 yrrAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100  
 7862 ACGTATTTTCCCATTTGATGCTCCGAGCAACGTTGCTTTGGTCTT 7813  
 101 MetGlnLysValProLysGluLeuIleGlnGlnThrAsnLysTyr 117  
 7812 ATGCAAAAGAAGCTCCAAAAGAAGATTGATTACAGACCAACAAGTA 7763  
 117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysL 134  
 7762 TCTTGAACTCATGCAAAATGCTCAATATGCGATCGAAAGCCCATTAAC 7713  
 134 euserGlyGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150  
 7712 TCAAGTGTGGACACACACGTCACCTTGGCATCGCCCTTAGCGGTT 7663  
 151 AspProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLys 167  
 7662 AATCCAACTGTCTCTCATGAGACGACCACTTGTATCTGGAGCCCA 7613  
 167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184  
 7612 ACTTCGCTTGATGCGTCACGACCATCCGAAATCCACACAGAAATGG 7563  
 184 LylIleThrThrValTyrValThrHisAspGlnGluGluAlaMetAlaIle 200  
 7562 GAATTACAACTGTTATGTATGTAACCCAGCACCAAGAAGAACCATGGTATT 7513  
 201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArg 217  
 7512 TCAGAACCAATTCGCTGTATGTAAGAATGGGGTGATCCACAAATCGGCCG 7463  
 217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234  
 7462 ACCNAAGAAGACTCTATCAACACGACTAATGAGTTGTGGCAACCTTTA 7413  
 234 leGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250  
 7412 TCGGACCCACCAAAATATATCCCTGCCAATCTGAAAAACGAGGAGCGGC 7363  
 251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267  
 7362 GCTTATATCGCTTTTCAGATGCGTATGCCCTTGCAATGCCAGCTTTGA 7313  
 267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluLubPheI 284  
 7312 TCAGGTGAGGAGCAAGCTATTCATGTAAGCATTCGTCGCAAGATTTA 7263  
 284 leysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300  
 7262 TCAAGAATGATCTGGAGATATTGAAGCACTATTGAAGATAGCGCTTAT 7213  
 301 leuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317  
 7212 CTTCGACTTAATACGATATTATTCATTGAGACAGAGTTTTCCTCAAAAT 7163  
 317 eGlnValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 334  
 7162 TCAAGTTAGGAAGATCAACTTTTGAAGAAGACTCAAAAAGCAATC 7113  
 334 rglIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350  
 7112 GTATTGTCTACGATCAATACGCAAAATTTAAACATCTTTTCGCGAGAT 7063  
 351 gLysSerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363  
 7062 GGTTCCTCCAAACCTGATATAAAGAGTCAACCATGCAACG 7024

seq\_name: gb\_ba:AE007337  
 seq\_documentation\_block:  
 LOCUS AE007337 10216 bp DNA linear BCT 31-Aug-2001  
 DEFINITION Streptococcus pneumoniae TIGR4 section 20 of 194 of the complete genome.  
 ACCESSION AE007337 AE005672  
 VERSION AE007337.1 GI:14971709  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Streptococcus pneumoniae TIGR4.  
 Streptococcus pneumoniae TIGR4  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.  
 1 (bases 1 to 10216)  
 Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,D.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
 Complete genome sequence of a virulent isolate of Streptococcus pneumoniae  
 Science. 293 (5529), 498-506 (2001)  
 2 (bases 1 to 10216)  
 Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,D.D., Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
 Direct Submission  
 Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Location/Qualifiers  
 1..10216  
 /organism="Streptococcus pneumoniae TIGR4"  
 /strain="TIGR4"  
 /db\_xref="taxon:170187"  
 153..845  
 /gene="SP0240"  
 153..845  
 /gene="SP0240"  
 /transl\_table=1  
 /product="phosphoglycerate mutase family protein"  
 /protein\_id="AAK74420.1"  
 /db\_xref="GI:14971710"  
 /translation="MKVRLVLRHGKTMFNTIGRAGCKSDPTLAREGRIOELGIC LRESLOFERAASSDSGRITDGMGILLELGIGETIPYRMDRIKREKPGSDGAYDGLDFMGIIIRIFNVHDVHQLSYAEIAGLEVDTAGAWGWEKLSGRIRKEGEMIAKEM EDOGGNALVYVSHGWTIGTIYVLINGMHPHGIDNSVTLIEVEDQFREVYVGDGSRYSR ELGRKMEGKST"  
 complement(948..2593)  
 /gene="SP0241"  
 /note="This region contains a gene with one or more premature stops or frameshifts, and is not the result of a sequencing artifact; This region contains an authentic frame shift and is not the result of a sequencing artifact.; similar to GP:276195; identified"  
 complement(2583..3674)  
 /gene="SP0242"  
 complement(2583..3674)  
 /gene="SP0242"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter, ATP-binding protein"

gene  
CDS

protein\_id="AAK74421.1"  
/translation="MSEIKIINAKRIYHDVPIENLITIPKSLFTLLGASGGCKTT  
/db\_xref="GI:14971711"  
LLRMFAGNSIEEGEFPDDTKINNMEPSKNIGMVONFAIFPHLTVRDVAFLMQ  
KVPEELIIOOTNTYELMOIAOYADRPKLSGGOOQRYALACALAVNVILMDPR  
LSNLPAKRLDMROAIRIRIOHENGITTYRVYHODEEMALSIDQJAWKKDVIOIGRP  
KELYHKPANEVVFATIGRTNITIPANLEKRSRGATIVSDGLAMPALDDVEBAIHV  
SIRPEEFIKDSSDIETGRDVSVALNDTFDIETGFASKIQVSESTFEEDLOKGNR  
IRLRINTOKLNIFSDAGSQNLKGVNHGT"  
complement(3687. .4703)  
/gene="SP0243"  
complement(3687. .4703)  
/gene="SP0243"  
/codon\_start=1  
/transl\_table=1  
/product="iron ABC transporter, iron-binding protein"  
/protein\_id="AAK74422.1"  
/db\_xref="GI:14971712"  
/translation="MKRWMTYAACSSESADSSSPKGDSGLSVYSPNSEGLIGAT  
IPAEEKAGIVEVLIOAGTGELFKLESEKVPYADVIFGGSYQYTTHGLEPENYS  
KENDVIAYEQONTGYSPYPLDGSVLVNDLPDKGMNIEGNDLFPELKGAJATAD  
PANSSAQAOLTNMLQAGGYKDKRAMSYVDLFDLDKIGSSSSSVKYVADGEWA  
VGSLSEDAVAVLLNDGANIKRIKVVYKEGVVELPASAIYKSKNMENAKRFIDEISOE  
VODLTGITTTNRPPVKRNKAKISENNKPIDIKITIEDDYIYKKSNDIYKRIINEVDI  
OSQ"

gene  
CDS

complement(5008. .5115)  
/gene="SP0244"  
complement(5008. .5115)  
/gene="SP0244"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAK74423.1"  
/db\_xref="GI:14971713"  
/translation="MSHSFKKSLQKEILLHRSSIAFYTSRAFSPTVSP1"  
complement(5290. .6066)  
/gene="SP0245"  
complement(5290. .6066)  
/gene="SP0245"  
/note="identified by match to Pfam protein family HMM  
PF02143"

gene  
CDS

/codon\_start=1  
/transl\_table=1  
/product="pyruvate formate-lyase-activating enzyme,  
putative"  
/protein\_id="AAK74424.1"  
/db\_xref="GI:14971714"  
/translation="MEISKGIIFNIQHSIHDPGIRTVTEFLKGCPLRCPCWCSNPESQ  
RKPKMKDAOREKFTLVGEERVAEITTEYLKAEFEYESGGGLTSGSIGVAFQFE  
AKAIPKSKHEHHIHAITEPAFVDERKITDLIOYVDFTIDKLHYNSIKKHXYGFMFN  
OMIKNIHYASONSQKIYLRIPVIPNFNSLEDNAEKRTPLENSLNIDQYQLLPFHOG  
ENKIYLNKRKTIMDSINALHPEDLDIOKVELNHHINC1"  
6188. .6934  
/gene="SP0246"  
6188. .6934  
/gene="SP0246"  
/codon\_start=1  
/transl\_table=1  
/product="transcriptional regulator, Deor family"  
/protein\_id="AAK74425.1"  
/db\_xref="GI:14971715"  
/translation="MKRLBOIIRKLYSEYEKIDVNTISEKLNVSVTIRKDCLKESNG  
LIHRHGAVYNSGDLDLVRLSIYNTERKKIYOEVAKYKVNIMTESSTCTLAAR  
CECKQRNAVITYTNSFTIANPYRAINDSCRVIYGEFGKDSQVTGFLKMIOTHTYR  
COAFGTGVDXDEMGFTGRDLMRSEVVOYISAVSDKIVTLIDSSKFDRKGIYRRFALS  
OYEVITBEKLSKOMIATLENAGIMWKVVS"  
6947. .7927  
/gene="SP0247"  
6947. .7927  
/gene="SP0247"  
/note="similar to SP:P37078 GB:X66059 PID:43938;  
identified by sequence similarity; putative"

/codon\_start=1  
 /transl\_table=1  
 /product="transcriptional regulator"  
 /protein\_id="AAK74426.1"  
 /db\_xref="GI:14971716"  
 /translation="MNDRRKLKSLKIAYLYIENLANSQIAAKIGIYRTSISMLTPEA  
 RNGIYKTEIENEDPTNMFKLENNYKKEYSLESLEITINEDDDIPTIISEKISQVACV  
 LRNLIDNNMIGFSWGSLSNLVDLHKSIVRANHFYPLAGDSHIAKHAVNTLLYE  
 MSRFHECEFTFMATLYOENKLLADGILOSRVEPNLKNWKDDIAVAGIGDFSNGK  
 HMDMLTEDEDFEKLTVEKTKYGEICRFPSKGEYENQERTIAISLEDLKNIPQS  
 LAVAGSTKVSLSISVLRANLVNHLIDKNITLKLEDEDDLPFRELIG"  
 8122..8442  
 /gene="Sp0248"  
 8122..8442  
 /gene="Sp0248"  
 /codon\_start=1  
 /transl\_table=1  
 /product="PTS system, IIA component"  
 /protein\_id="AAK74427.1"  
 /db\_xref="GI:14971717"  
 /translation="MEMIVPQILIMGLILYAGDAKQHIYKALDYIKNGTCERCEEEIQ  
 LADALAEAHNLQTKFLAQSASGKTETFLAEVSHQHLMTSMTEINLKEIISLKEE  
 LHKY"  
 8488..8796  
 /gene="Sp0249"  
 8488..8796  
 /gene="Sp0249"  
 /codon\_start=1  
 /transl\_table=1  
 /product="PTS system, IIB component"  
 /protein\_id="AAK74428.1"  
 /db\_xref="GI:14971718"  
 /translation="MVKIGLFCAGSFTGMLVNMMKIAAOSGVEAEIEAFSOSKLAD  
 YAPNIDVALLGPQVAYFLDKSKEICDCDVP IAVIPMDYGMDCRKLVDLALSLSG  
 "

8789..10111  
 /gene="Sp0250"  
 8789..10111  
 /gene="Sp0250"  
 /codon\_start=1  
 /transl\_table=1  
 /product="PTS system, IIC component"  
 /protein\_id="AAK74429.1"  
 /db\_xref="GI:14971719"  
 /translation="MGKKRRFIMSKMDVOKIIPAMKFNVMRGIIALKDGLMALTPLT  
 VSGSLFIMQLPREGINKLSIASVEGNAMPPEMOVSGEFLMGLISCSIASVYAK  
 NSGEALPAGVLSAFETILKRSYIRKQGEALIDAIKSWFGQGITGAILIELVYG  
 SIYEFITKRKIVIMPEQVPOALIKOPEAMIPAVEILSVIYLAKSLNGSTFLE  
 MIAVEIYDGLTGLSGLYGAIGIAFISFLMWFECVHOSVNGVATLLSLNDPANKAA  
 MLANSLIENGCAHIVQOFLDSFLISGTFISGLVAMFAKSKOYOLGKVAAP  
 PAIVNBPVYFGEFPIVNMVPMFVPLIIVPLAVIYGAIAATGFMQFSGVTLPMST  
 PAISGLFVGMGCVITQVLIAMSTLVYFFPFVQDRALVQNEIKQS"

BASE COUNT 3252 a 1519 c 2088 g 3357 t  
 ORIGIN

alignment\_scores:  
 Quality: 1843.00 Length: 363  
 Ratio: 5.077 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-769-787-162 x AE007337/rev ..

Align seg 1/1 to reverse of: AE007337 from: 1 to: 10216

1 MetSerGluIleLysIleIleAsnAlaLysLysIleTyHISpValPr 17  
 |||||  
 3674 ATGAGTGAGATCAAAATTATTACGCCCAAAAAAATGTACACAGATGTCC 3625  
 |||||  
 17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34  
 |||||  
 3624 TGTATTATGGAATTTGACATTTCCAAAGAGAGATCTCTTTACCC 3575

34 euleuglyalaSerGlyCysGlyThrThrLeuLeuArgMetIleAla 50  
 3574 TTCTTGGAGCTTCAGATGTGGGAAAAAGACCTTCTTCGATGATGCA 3525  
 51 GlypheasnsrIleGlyGlyGluPheThrPheAspAspThrIle 67  
 3524 GGTTCACAGTATCGAAGGTGGAGAAATTTACTTCATGATACAAAAAT 3475  
 67 easnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84  
 3474 CAATAAATATGGAACCAAGCAACGCAATATGGATGTTTCCAAAAC 3425  
 84 yralaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100  
 3424 ACGTATATTTCCCATTTGACTGTCCGAGCAACGTTGCTTTGGTCT 3375  
 101 MetGlnLysLysValProLysGluGluLeuIleGlnGlnThrAsnLys 117  
 3374 ATGCAAAAGAGGTTCCAAAAGAAATGATTCACACAGCAACAAGTA 3325  
 117 rleugluleuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLys 134  
 3324 TCTTGAACCTCATGCAATTCGCTCAATATGGGATCGAAAGCCGATTAAC 3275  
 134 euSerGlyGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150  
 3274 TCAGTGTGGACACACAAACGTCACCTGGCATCGCGCTTGGCGTT 3225  
 151 AsnProSerValLeuLeuMetAspGluProLeuSerLysLeuAlaLys 167  
 3224 AATCCAAGTGTCTCCATGAGGAGCCACTAGTATCTGGGAGGCCAA 3175  
 167 sleuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184  
 3174 ACTTCGCTTGATATGCGTCAAGCATCCGAAATCCAAACACAGAAATGG 3125  
 184 lyllethrThrValTyrValThrHisAspGlnGlnGluAlaMetAlaIle 200  
 3124 GAATTACAACTGTTATGTAAACCACGACCAAGAAAGCAACCATGCTATT 3075  
 201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGly 217  
 3074 TCGAACCAAAATGCTGTTATGAAGAGGGGTGATCCAAACAAATGGCGC 3025  
 217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234  
 3024 ACCAAAGAACTCATATACCAAGCTAATGAGTTTGTGGCAACCTTTA 2975  
 234 leglYargThrAsnIleIleProAlaAsnLeuGlyLysArgSerAspGly 250  
 2974 TCGGACCCCAAAATATATATCCCTGCCAATCTTGAAGAAACGAGCAGCGC 2925  
 251 AlATyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267  
 2924 GCTTATATCTCTTTTCAGATGGCTATGCCCTTCGAATGCGACGCTTGA 2875  
 267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluGluPhe 284  
 2874 TCAGGTTGAGAGCAACGCTATTCATGAAACATTCGTCGCCGAGAGTTTA 2825  
 284 leLysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300  
 2824 TCAAGATGATCTGAGATATGAAAGAACTATTAGAGATAGCGTCTAT 2775  
 302 leuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysIle 317  
 2774 CTTGGAGCTAAATACGATTAATTTTCATTGAGACAGGTTTCCCTCAAAAT 2725  
 317 eGlnValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 334  
 2724 TCAAGTTAGTGAAGATCAACTTTTGAGAGAAGATCTTCAAAAAGGCATC 2675

334 rgllleargLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350  
 2674 GTATTCCTTACGAAATCAATACGCAAAATTAACATCTTCTCGCAGAT 2625  
 351 GlySerGlnAsnLeuIleLysGlyValAlaSerHisGlyThr 363  
 2624 GGTTCCTCAAAACCTGATTAAGAGATCAACCATGAGACG 2586

seq\_name: gb\_ba:AE008403

seq\_documentation\_block:

LOCUS AE008403 9976 bp DNA linear BCT 13-SEP-2001

DEFINITION Streptococcus pneumoniae R6 section 19 of 184 of the complete

genome.  
 accession AE008403 AE007317  
 version AE008403.1 GI:15457762

KEYWORDS

SOURCE

ORGANISM

Streptococcus pneumoniae R6.  
 Streptococcus pneumoniae R6  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.

REFERENCE  
 1 (bases 1 to 9976)

AUTHORS

Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,  
 Dehoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Gerlinger, C.,  
 Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Ladace, R.,  
 Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,  
 Mcahren, S., McHenry, M., Mcleaster, K., Mundy, C., Nicas, T.I.,  
 Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P.,  
 Sun, P.-M., Winkler, M.E., Yang, Y., Young, Bellido, M., Zhao, G.,  
 Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L.  
 and Glass, J.I.  
 Genome of the Bacterium Streptococcus pneumoniae Strain R6  
 J. Bacteriol. 183 (19), 5709-5717 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 9976)  
 Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,  
 Dehoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Gerlinger, C.,  
 Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Ladace, R.,  
 Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,  
 Mcahren, S., McHenry, M., Mcleaster, K., Mundy, C., Nicas, T.I.,  
 Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P.,  
 Sun, P.-M., Winkler, M.E., Yang, Y., Young, Bellido, M., Zhao, G.,  
 Zook, C., Baltz, R.H., Jaskunas, S.R., Rostock, P.R. Jr., Skatrud, P.L.  
 and Glass, J.I.  
 Direct Submission  
 Submitted (27-JUN-2001) Infectious Diseases Research, Eli Lilly and  
 Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Location/Qualifiers  
 1..9976  
 /organism="Streptococcus pneumoniae R6"  
 /strain="R6"  
 /db\_xref="taxon:171101"  
 117..476  
 /gene="spr0217"  
 117..476  
 /gene="spr0217"  
 /note="ACT domain containing protein"  
 /codon\_start=1  
 /transl\_table=1  
 /label="spr0217"  
 /product="Conserved hypothetical protein"  
 /protein\_id="AAK9021.1"  
 /db\_xref="GI:15457763"  
 /translation="MSIFVRLMYIKRIENVAVSEFKKMSILMKAITVVGKDS  
 GIVAGVSKIELGLNIDDISQTVLDEFTFMVAVSSDEKODFTYLNREFAFGQTIN  
 VKINQSAIPEAMYN"

gene

CDS

486..1823  
 /gene="spr0218"  
 486..1823  
 /gene="spr0218"  
 /codon\_start=1  
 /transl\_table=1

gene

CDS

486..1823  
 /gene="spr0218"  
 486..1823  
 /gene="spr0218"  
 /codon\_start=1  
 /transl\_table=1

gene

CDS

486..1823  
 /gene="spr0218"  
 486..1823  
 /gene="spr0218"  
 /codon\_start=1  
 /transl\_table=1

/label=spr0218  
 /product="Conserved hypothetical protein"  
 /protein\_id="AAK9022.1"  
 /db\_xref="GI:15457764"  
 /translation="MDIROVETETAMIEONFDIRTYMGISLDCIDPNNINRAEKI  
 YOKITTKTAAVAGDEIAEAELFPIYKKRVSVPISLIGAAPATQYVLAALADKA  
 AKELGVVDITIGFSALVQKGYOKGDEIILNSTPRALAEIDKCSSVNIIGTSKINMTA  
 VADKRIILIKETANLSDMGVAKLVVFAANVEDNPPMAGFHVGGADVIINGVSGPQV  
 VKRALEKFGSGFDVVAETVTKAFKIRIQOLVGOMASERLGEVFIIVDSLPTPA  
 VGDVAVRLYEELGSETVGTGHTTALAALLNPKVKGGMACNOYGGSLGAFIPSEDEE  
 CHIAVNGNSLLEKLEAMTAICSGIDMIVIPDTPRETIAAMIADEAIGVINMT  
 TAVRIIPFGREGDEDMEEFGGLLTGAPVMKVNAGASSVDFTISGQILPITHSKN"  
 2010..2759  
 /gene="gpmb"  
 /note="spr0219"  
 2010..2759  
 /gene="gpmb"  
 /EC\_number="5.4.2.1"  
 /codon\_start=1  
 /transl\_table=11  
 /label=spr0219  
 /product="Phosphoglycerate mutase 2 paralog"  
 /protein\_id="AAK9023.1"  
 /db\_xref="GI:15457765"  
 /translation="MGGFYFIIDHIEANGSRKWKVRLIYRHKTFMNTIGNAGCM  
 SDPTPLTAGSERIGLELIGLRESDDQFERAYSDSGRTIQMGIIIEELGQGIPIR  
 MDKIRREMGFSGFDGAYGDDLFMGIIPRIFNVDVHOLSVALEGLVEDYPAAGMAG  
 LEYELSGRIKREGFEMIAKEMEDOGGNAIYVSHGMTIGTYILINGMHPHGDNSVTI  
 LEYEDGQFRRVEVAGDGRSYRELGRKMEBGS"  
 complement(2862)..3896  
 /gene="ABC-MSP"  
 /note="spr0220"  
 complement(2862)..3896  
 /gene="ABC-MSP"  
 /note="Putative"  
 /codon\_start=1  
 /transl\_table=11  
 /label=spr0220  
 /product="ABC transporter membrane-spanning permease -  
 iron transport"  
 /protein\_id="AAK9024.1"  
 /db\_xref="GI:15457766"  
 /translation="MLVFRAFSDGFTPLMIGEGRTFPPVIITYQFISENGNSAFAI  
 MAIIIALIFLOKHIANRYSFNSMLHPIDPKTKGKMAIYATVYGIIFISVLPQ  
 ILYLTSFELKTSGMVEFGFGYSPNSYKLAIFNMGAIFENTIRIPLIALVVLVAFPTFS  
 YLAVERKRLNFETLIDLSIMWPYIVPGTVLGIATFISFEMTGLFGSGPLMTGTATILIM  
 SLAKRLIYTRIRSSVASIQQIAPSTEEAASIGSSRLNIFAKITPPMILSITIGAILIL  
 SMVTSILSELSTILYINVKTRMYAITYTEVLRNGYGAALSTILYLVYGSLLFM  
 KISKNSNITL"  
 complement(3881)..4507  
 /gene="ABC-MSP"  
 /note="spr0221"  
 complement(3881)..4507  
 /gene="ABC-MSP"  
 /note="Putative"  
 /codon\_start=1  
 /transl\_table=11  
 /label=spr0221  
 /product="ABC transporter membrane-spanning permease -  
 iron transport"  
 /protein\_id="AAK9025.1"  
 /db\_xref="GI:15457767"  
 /translation="MERRKLNMTASSFEFLTYLVLPVPIITVLKQALIHGQFSL  
 ANFVPEFGAYVSEVLNVSFRYSITATVTSILVGLIAGFSGMDPGGKRLQIILITII  
 ASMSAPFGASVSVLLLRGNVITKEFLNALHPLAIDIGKGIYVLPITQLPFLVFL  
 VYAGTWNISDNLSEPAASMGSGFGKPLVTVVLPVLFTLLAACPLYL"  
 complement(4497)..5588  
 /gene="ABC-NBD"  
 /note="spr0222"  
 complement(4497)..5588  
 /gene="ABC-NBD"  
 /note="Putative"  
 /codon\_start=1

	/transl_table=11	
	/label-spr0222	
	/product="ABC transporter ATP-binding protein - iron	
	transport"	
	/protein_id="AAK9026.1"	
	/db_xref="GI:15457768"	
	/translation="MSIEIILNAKKIYHDVPIENLNTIPKGSFLPLGASCGRTT	
	LNRKIAGFNSIKDGEFFEDPTKINNEPSPKRNLGVONATIFPHLVIRNVAFGLMO	
	KPKVEKLIDQTNKYLYELMOIAQVADRPKRLSGGOQRVTALACALAVNSVLMEP	
	LSNLEAKRLDMROTIREIOHEVGITIVYTVDHOEEAMASDOIAVMKDSVIGQISRP	
	KEYLHKRANFEVAATFIGRTNIIPANLEKRSDDGAYIVPSDGVALRMPAKDVEOAHV	
	SIREETFKDESGDIESTIRDSYGLNTGYFIETGASKIQVSEESTFBEELQKGRR	
	IRLRINTQKLNTSPADSQNLIKVNHG"	
gene	complement(5601..5969)	
	/gene="ABC-SBP-truncation"	
	/note="spr0223"	
CDS	complement(5601..5969)	
	/gene="ABC-SBP-truncation"	
	/note="Putative"	
	/codon_start=1	
	/transl_table=11	
	/label-spr0223	
	/product="ABC transporter solute-binding protein - iron	
	transport_truncation"	
	/protein_id="AAK9027.1"	
	/db_xref="GI:15457769"	
	/translation="MTVGLSYEDSAVKLLNDGANIKVVYPKEGTVELPASAAIVKSK	
	NMEAKFIDFIISOEVDPILGTTTTRPVRRNAKTSSENMRPIDIKLTLEDYDVYIK	
	NKSIVKYKNVEVFSDIQSQ"	
repeat_region	6020..6126	
	/rpt_family="RUP element"	
	/rpt_type-dispersed	
gene	/rpt_unit-Rup_A	
	complement(6114..6308)	
	/gene="ABC-SBP-truncation"	
	/note="spr0224"	
CDS	complement(6114..6308)	
	/gene="ABC-SBP-truncation"	
	/note="Putative"	
	/codon_start=1	
	/transl_table=11	
	/label-spr0224	
	/product="ABC transporter substrate-binding protein - iron	
	transport_truncation"	
	/protein_id="AAK9028.1"	
	/db_xref="GI:15457770"	
	/translation="MNTEGSDLKLPCLKGIATPADNSSAPAQLTNNMLQAQGCK	
	DDRKAISTVADLFILIDICVK"	
repeat_region	6799..6897	
	/rpt_family="RUP element"	
	/rpt_type-dispersed	
gene	/rpt_unit-Rup_B1	
	complement(6890..7137)	
	/gene="spr0225"	
CDS	complement(6940..7137)	
	/gene="spr0225"	
	/codon_start=1	
	/transl_table=11	
	/label-spr0225	
	/product="Hypothetical protein"	
	/protein_id="AAK9029.1"	
	/db_xref="GI:15457771"	
	/translation="MTLNEXYLQYRLKQALDKMAESPNSPLAISDQVGSDDYKFPAK	
	VFKYHLISPRKSLGRIVK"	
gene	complement(7212..7988)	
alignment_scores:		
Quality:	1817.00	Length: 363
Ratio:	5.019	Gaps: 0
Percent Similarity:	99.725	Percent Identity: 98.347
alignment_block:		
US-09-769-787-162 x AE008403/rev ..		

Align seg 1/1 to reverse of: AE008403 from: 1 to: 9976

```

1 MetSerGIuIeLysIleIleAsnAlaLysLysIleTyRHisAspValPr 17
|||||
5588 ATGAGTGCATCGAATATTATTAACCCCAAAATAATCTACCGAGATGTC 5539
17 OvaIIeGIuAsnLeuAsnIleThrIleProLysGIySerLeuPheThrL 34
|||||
5538 TGTATTGAGATTGGAACATTACAAATCCAAAGAGAGTCTCTTACC 5489
34 euleuGIyAlaSerGIyCysGIyLysThrIleLeuAlaArgMetIleAla 50
|||||
5488 TTCCTGGAGCTTCAGAGATGGGAAACGACCTCTTCGATGATCA 5439
51 GlyPheAsnSerIleGIuGIyGIyLupPheTyRAspPheThrLysI 67
|||||
5438 GCTTCAACAGATCAAAAGATGAGAAATTTACTTCGATGATCAAAAAT 5389
67 eAsnAsnMetGIuProSerLysArgAsnIleGIyMetValPheGlnAsn 84
|||||
5388 CAATTAATATGGAACCCAGCAAAATATCGGCTGTTTCCAAACT 5339
84 yRAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGIyLeu 100
|||||
5338 ACGTATTTCCTCCACATTGACTGTCCAGACAACTGCTTTGGTCTT 5289
101 MetGIyLysValProLysGIuGIuLeuIleGlnIleThrAsnLysTy 117
|||||
5288 ATGCCAAAGAAAGCTTCCAAAGAAAGATTTGATTCACAGACCAACAATA 5239
117 rLeuGIuLeuMetGlnIleAlaGlnTyRAlaAspArgLysProAspLysL 134
|||||
5238 TCTTGAACCTCAATGCAAAATTCCTCATATGCGGATCGAAGCCGATAAC 5189
134 eusSerGIyGIyGlnGlnIleArgValThrIleAlaCysAlaLeuAlaVal 150
|||||
5188 TCAGTGGGAGCAACACACGCTGCACCTGGCTGCGCTTACCGCTT 5139
151 AsnProSerValLeuLeuMetAspGIuProLeuSerAsnLeuGlnAlaL 167
|||||
5138 AATCCAAATGTTCTCTCATGAGACGACCTAGTATCTGGAGGCCAA 5089
167 sLeuArgLeuAspMetArgIleAlaIleArgGluIleGlnHisGluValG 184
|||||
5088 ACTTCGCTTGATATGCTCAAACTCCGAGAAATCCAAACAGCAAGTGG 5039
184 LyIleThrThrValTyRValThrHisAspGlnGIuAlaMetAlaIle 200
|||||
5038 GAATTCAACTGTTTATGTAAACCCAGCACCAAGAAAGACCATGGCTATT 4989
201 SerAspGlnIleAlaValMetLysAspGIyAlaIleGlnIleGIyArg 217
|||||
4988 TCAGACCAAAATTCGCTTATGAAGATGGGTGATCCAAACAATCGCGCG 4939
217 gProLysGIuLeuTyRHisLysProAlaAsnGIuPheValAlaThrPheI 234
|||||
4938 ACCAAAGAACTCTATCATTAACACAGCTAATGAGTTTGGCAACTTTTA 4889
234 leGIyArgThrAsnIleIleProAlaAsnLeuGIuLysArgSerAspGIy 250
|||||
4888 TCGGACGCGCAAAATATTATCCCTGCCAATCTTGAAGAAAGGAGCAGCGC 4839
251 AlaTyRileValPheSerAspGIyTyRAlaLeuArgMetProAlaLeuAs 267
|||||
4838 GCTTATATCGCTTTTCAGATGGCTATGCCCTTCGAATGCCAGCTCTTGA 4789
267 gGIuVgIleGIuGIuAlaIleHisValSerIleArgProGIuGIuPheI 284
|||||
4788 TCAGTTGAGAGCAAGCTATTATCATGTAAGCATTCGTCCGAAGAGTTTA 4739
284 leLysAspGIuSerGIyAspIleGIuGIyThrIleArgAspSerValTyR 300
|||||

```

```

4738 TCAAAGATGAATCTGCAGATATCGAAGGAACTATTAGATAGCGCTTAT 4689
301 LeuGIyLeuAsnThrAspTyR-PheIleGIuThrGIyPheAlaSerLysII 317
|||||
4688 CTGGACTTAATATACGATATTATTTCTATGACAGGTTTGGCTCAAAAAT 4639
317 eGlnValSerGIuGIuSerThrPheGIuGIuAspLeuGIuLysGIyAsnA 334
|||||
4638 TCAAGTTAGTAGAAGATCACTTTTGAAGAGAACTACAAAGAGCAATC 4589
334 rGIIeArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
|||||
4588 GTATTGCTTACGAAATCAATACGCAAAATAATTAACATCTTTCTGCAGAT 4539
351 GlySerGlnAsnLeuIleLysGIyValAsnHisGIyThr 363
|||||
4538 GGTTCGCCAAACCTGATTAAGAGACTCAACCATGGAACG 4500
seq_name: gb_hlg:SPNEU1908
seq_documentation_block:
LOCUS SPNEU1908 175936 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
ACCESSION AL449930
VERSION AL449930.1 GI:11545155
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Streptococcus pneumoniae.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
REFERENCE
AUTHORS 1 (bases 1 to 175936)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
2 (bases 1 to 175936)
Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
and Garcia-Bustos,J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1.175936
/organism="Streptococcus pneumoniae"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 52994 a 31655 c 38793 g 52447 t 47 others
ORIGIN
alignment_scores:
Quality: 1807.00 Length: 363
Ratio: 5.006 Gaps: 0
Percent Similarity: 99.449 Percent Identity: 97.796
alignment_block:
US-09-769-787-162 x SPNEU1908/rev ..
Align seg 1/1 to reverse of: SPNEU1908 from: 1 to: 175936
1 MetSerGIuIeLysIleIleAsnAlaLysLysIleTyRHisAspValPr 17

```

```

|||||.....|
69746 ATGATGAGATCGAAATTTAATTAACGCCAAAAAATCTACACATGTCGCC 69697
47 ovalleigliuasnleuasnleuThrleProlysglyserleuPheThrL 34
|||||.....|
69696 TGTATTATGAAATTTGAACTTCAATTCACAAAAGAAAGCTCTGTTACCC 69647
34 euleuigliuAlaSerGlyCysGlyLysThrThreuleuArgetleuAla 50
|||||.....|
69646 TTCTTGAGACTTCAGATGTGGAAAAAGACCCCTTCTGCTATGATGCA 69597
51 GlyPheAsnSerlleigluclygluPheTyrPheAspAspThrLysI 67
|||||.....|
69596 GGTTCACAGATCAAGATGAGAAATTTACTGATGATGATCAAAAAT 69547
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsp 84
|||||.....|
69546 CAATATATATGAAACCAACCAACGCAATATGGGGTGTGTTTCCAAAAC 69497
84 yAlaIlePheProHsleuThrValArgAspAsnValAlaPheGlyLeu 100
|||||.....|
69496 AGCCTATTTTCCACATTTGACGTCCGAGACACAGTTGCTTTGGTCTT 69447
101 MetGlnLysLysValProLysGluLysLysIleGlnThrAsnLysTy 117
|||||.....|
69446 ATGCAAAAGAGGTTCACAAAGAAATGATTCACACAGCAACAAGTA 69397
117 leuGluLysMetGlnIleAlaGlnLysLysLysLysLysLysLysLys 134
|||||.....|
69396 TCTTGAACATGCAAAATGCTCAATATGCGGATGCAAAAGCCCGATAAC 69347
134 euserGlyGlnGlnGlnArgValThrleuAlaCysAlaLeuAlaVal 150
|||||.....|
69346 TCAGTGTGTGACACAAACACGTCACCTTGGCATGCGCTTAAGCGTT 69297
151 AsnProSerValleuLeuMetAspGluProLeuSerAsnLeuGluLys 167
|||||.....|
69296 AATCCAAAGTGTCTCTCATGAGAGCCACTAGTATCTGAGGCCAA 69247
167 sleuAlaGleuAspMetArgGlnAlaIleArgGluIleGlnHsGluValG 184
|||||.....|
69246 ACTTCCCTTGATATGCTCAAGCCATCCGAGAAATCCAAACAGAGTGG 69197
184 lyIleThrValIleValIleThrHisAspGlnGluLysAlaIle 200
|||||.....|
69196 GAATTCACACTGTTTATGTAACCCACGACCAAGACAGCAAGCTATT 69147
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyLys 217
|||||.....|
69146 TCAGACCAATTCGTGTATGAAAGATGGGATGATCCAACAATCGGCC 69097
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||.....|
69096 ACCAAAGAACTCTATCATTAACACACTATGATGTTGGCAACCTTTA 69047
234 leGlyArgThrAsnIleleProAlaAsnLeuGluLysArgSerAspGly 250
|||||.....|
69046 TCGGAGCGCAAAATATTATCCCTACTAATCTTAAAAAAGCGAGCGCGC 68997
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
|||||.....|
68996 GCTTATATCGTCTTTCAGATGCTATGCCCTTCGATGCGCACTTCTGA 68947
267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluGluPhe 284
|||||.....|
68946 TCAGGTGGAGGAGCAAGCTATTCATGTAGCATTCCTCCGGAAGCTTTA 68897
284 leLysAspGluSerGlyAspIleGlyLysLysLysLysLysLysLysLys 300
|||||.....|
68896 TCAAAATGATCTCGAGATATTGAAGGAACCTATTAGCGATACGCTAT 68847
301 leuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317
|||||.....|
68846 CTTCGACTAATAATCAGGATATTTCAATTGACACAGCTTTGCTCCAAAAT 68797
317 eGluValSerGluGluSerThrPheGluGluAsnLeuGlnLysGlyAsnA 334
|||||.....|
68796 TCAAGTTAGTGAAGATCAACTTTTGAAGAGATCTTCAAAAAGGCAATC 68747
334 rgIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
|||||.....|
68746 GTATTGCTTCAGCAATCAATACGCAAAAATTAAACATCTTTCTGAGAT 68697
351 GlySerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363
|||||.....|
68696 GGTTCACAAACCTGATTAAGAGATCAACCATGAAAGC 68658
seq_name: gp_ba:SHU75349
seq_documentation_block:
LOCUS SHU75349 6200 bp DNA linear BCT 18-NOV-1999
DEFINITION Serpulina hydrosentariae bit operon, complete sequence.
ACCESSION U75349
VERSION U75349.3 GI:6127224
KEYWORDS
SOURCE
ORGANISM
Brachyspira hyodysenteriae.
Brachyspira hyodysenteriae
Bacteria; Spirochaetales; Brachyspiraceae; Brachyspira.
1 (bases 1 to 6200)
Dugourd,D., Martin,C., Rioux,C.R., Jacques,M. and Harel,J.
Characterization of a periplasmic ATP-binding cassette iron import
system of Brachyspira (Serpulina) hyodysenteriae
J. Bacteriol. 181 (22), 6948-6957 (1999)
JOURNAL
MEDLINE
20026807
PUBMED
10559160
2 (bases 1084 to 6183)
Dugourd,D.F., Jacques,M. and Harel,J.
Direct Submission
Submitted (17-OCT-1996) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
3 (bases 601 to 6200)
Dugourd,D.F., Jacques,M. and Harel,J.
Direct Submission
Submitted (05-MAR-1999) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
REMARK
Sequence update by submitter
4 (bases 1 to 6200)
Dugourd,D.F.
Direct Submission
Submitted (27-OCT-1999) Pathology/Microbiology, University of
Montreal, 3200 Sicotte, C. P. 5000, St-Hyacinthe, Quebec J2S 7C6,
Canada
REMARK
Sequence update by submitter
On Oct 27, 1999 this sequence version replaced gi:4337124.
COMMENT
FEATURES
source
1..6200
/organism="Brachyspira hyodysenteriae"
/db_xref="taxon:159"
119..6140
/gene="bit operon"
119..1129
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="periplasmic-iron-binding protein Bita"
/protein_id="AADI8094.2"
/db_xref="gi:6127225"
translation="MKNPILICMILMSWTLFSCSSGDSKMANSLVLYCSHPDLMMT
ILDFPKAKNPIDINEVYAGTGLIKRVARAKMNPIDGVLMGTLNSYKSKTDFENY
TSTNEANTLDFEKNTBEPTFESAIPSLIMNTNLAKIKIEGEDILNPKRTIAA
ADPSNASSAFENLVMILAMKRGPEKMDVVOALCAULDKLISGSAYKGVADGS
TVGLITKEEPGISYMSGSPVKVLYMEGVISKPDGVYIIGKKNLNAKKEIDYCVS
LDAQNMVLVEKLSRISIRSDAVVTDMVMKPMSEIYSITPNADVVEESROKMDKFDIFT
SI"

```

```

CDS
1164..2192
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="periplasmic-iron-binding protein BitB protein"
/protein_id="AAB95370.1"
/db_xref="GI:2766192"
/translacion="MKKILICSLVFLSLALFYSCSSKESQSGOAGNSLYICPHP
LEFINPLVDDEKAKNGINVDIIAGVGLIKRVESEKDNPLGDIILMGSLTKVPRK
ELEFNTTNEASIIADAYNKGAIREFALPSIVNNTNLANGIKIEGEDLLNDEL
KGKIAFDPSASSSEFHLVNMVAMNGDPGMDYISKLCANDGKILSGSSAVYK
GVADSEYTYGLTFEBEGGANYVSAQSPVKLVYKEGVIIKPDGITYIKNAKKNENAKK
VDATSYDAQKTIINDKLNRRSVRNDLPPSAILQSVDTINVITDDEAVDQNKMLNKK
EKDIFTSI"
2274..3293
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="periplasmic-iron-binding protein BitC"
/protein_id="AAB95371.1"
/db_xref="GI:2766193"
/translacion="MKKIVLFTSLLSVFIYSCSSSESGAOSGNSLYICPHPLEF
INPLVDDEKAKNGINVDIIAGVGLIKRVESEKDNPLGDIILMGSLTKVPRK
ESYSTNEENIAEIKNTGALTCTCTAPSLIMVNTNLAGDIKISYEDLLNDELKKA
IAFADPSASSSEFHLVNMVAMNGDPGMDYISKLCANDGKILSGSSAVYK
GVADSEYTYGLTFEBEGGANYVSAQSPVKLVYKEGVIIKPDGITYIKNAKKNENAKK
VDATSYDAQKTIINDKLNRRSVRNDLPPSAILQSVDTINVITDDEAVDQNKMLNKK
EKDIFTSI"
3375..4499
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="putative ABC transporter BitD"
/protein_id="AAB95372.1"
/db_xref="GI:2766194"
/translacion="MSVSISINVYKRYEKLITPDLSIEINGEFTLIGSGCGKT
TLRLMAGNTIEGEIKFDKDVINNTIPAKKNIGAVENVAIFPHMVRREVEGLK
LRKENKESKKRVDEMLHVKVIEYODRLPERLSGOQOVALAIAVITPVLAMDSE
PLSNLDAKLRIMERSAIDIKRHVGITIVYVHDEQEEALADRIAVMNGVYIOQVGS
PVSITRPNVAVATFIGHNSLFVATIKIEGDYTLFPGCGKLMKMDLNDKDDDEV
VYGIREEFVENSEDEGIIKAKILSKTFEGKTNFLEHNDNEVVPDPSIIEYSDSS
YTDPMTERDEVITLKPNAKINVTTPDMKSLIKGVKKE"
4492..3525
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="putative permease BitE"
/protein_id="AAB95373.1"
/db_xref="GI:2766195"
/translacion="MNKRLKIDFWYITLIIAIVAFALFLVPLFSLFISSEKDPETGA
WTLDNFLREFSRKYYOALVNSFVITCVTLIAIYIGPMAYFMSYKIKGKTFLEIL
IISMMSPAFICAVSMILLGRSGVSTPEFAGFEGGLVFEGLVETIMPEYII
FMVYSGALSKIDVSLIMEAESIGCSPEKRVVITAMPILITLIGSLVFNMAALDFG
TPMLIGEGRTMPMYISEFVSEGVGNANFSMASINVTIAMPFIQAKYIVVKNHNL
L"
5367..6140
/gene="bit operon"
/codon_start=1
/transl_table=11
/product="putative permease BitF"
/protein_id="AAB95374.1"
/db_xref="GI:2766196"
/translacion="MSVVIHFIYFVFLSLIPOLTVIYTSFLLKTNRAFLVKGFSINS
YISFISLSTSTKNTYIGIITFIIVIMGMFIAYISPRKNILSVIDITMPEYII
PESVIGITILLAKNKPPIILSGTSMIISLVIRKMPYILKRSSALIVQISIMEAS
ISLGSQVKEFEKIIAMMMLPGVISGALISWTVINESSVILVTGSRMAVSIYQ
EVIRASYSYGAALSTILSTIISLIEFKLKGKEVSL"
BASE COUNT      * 2278 a      788 c      1031 g      2102 t      1 others
ORIGIN
alignment_scores:
Quality: 919.00      Length: 366

```

```

Ratio: 3.137      Gaps: 4
Percent Similarity: 80.055      Percent Identity: 51.366
alignment block:
US-09-769-787-162 x SHU75349      ..
Align seg 1/1 to: SHU75349 from: 1 to: 6200
4  I l e y s i l l e a s n a l a t s y l s i e t y r h i s a s p v a l p r o v a l l e g l 20
:::|||||  |||  |||:  ::  |||:  ::  |||:  ::  |||:  ::  |||:  ::  |||:  ::  |||:  ::
3387  A T A C T T T G A A A A T G A T A A A A C G T T A T G A A A A G T T G A C A T A A T A C 3436
20  u a s n l e u a s n i l e t h r i l p r o l y s g l y s e r l e u p h e t h r l e u e n g l y a 37
:::|||||:  |||  |||  |||:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3437  T G A T C T T T C A T T A G A G A T A A A A A T G A G A A T T T T T A C T C T T G A C 3486
37  l a s e r g l y c y s g l y t h r t h r l e u e n u r g m e t l l e a l a g l y p h e a s n 53
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3487  C A T C A G A T C G G T A A A A C A C A C T C T T A C G T A T G A T T G C G G A T T T A A T 3536
54  s e r l l e u g l y g l y g l u p h e t y r p h e a s p a s p t h r l y s i l e a s n a s m e 70
:::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3537  A C A T A G A A G C G G A G A A A T A A A T T G A T A A G A T G A T A T C A A C A T A T 3586
70  t g l u p r o s e r l y s a r g a s n i l e g l y m e t v a l p h e g l a s n t y r a l a l e p 87
:::|||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3587  T C C T G C A C A T A A A A A A A T A T T G T A T G T A T T C A A A C A T A T G C A T A T 3636
87  h e p r o h i s l e u t h r v a l a r g a s p a s n v a l a p h e g l y l e u m e t g l n l y s 103
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3637  T T C C G A T A T A G A C T T T A G A A A A T G T T A A T A C G A T T A A A A C T C A G A 3686
104  l y s v a l p r o l y s g l u g l u e u l l e g l i n g l n t h r a s n l y s t y r l e u g l u e 120
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3687  A A G A A A A T A A A G A T C T A T G A G A A A A A A G T A G A T A A A T G C T C A T G T 3736
120  u m e t g l i l l e a l a g l i n t y r a l a a s p a r g l y p r o a s p l y s l e u s e r g l y 137
:::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3737  T G T A A A A T T G A A A A T A T A C A G C A C A C T C C G A A A G A T T A T C C G C G A G 3786
137  l y g i n g l i n a r g v a l t h r l e u a l a c y s a l a l e u a l a v a l a s p r o s e r 153
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3787  G A C A G C A G A G A G A G T T A G C A A G A C A A T A G T A T A C A C C A C A G T 3836
154  v a l l e u d e u e t a s p g l u p r o l e u s e r a s n l e u g l u a l a l y s l e u a r g l e 170
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3837  G T T T A C T A T T A T G A T G A G C C T T T A C T A C T A G A T G A T T A A T A A C A A T 3886
170  u a s p m e t a r g l a l a l l e a r g l u i l e g l i n h i s g l u v a l g l y l e t h r t 187
:::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3887  A G A A A T G A A G A C C A A T A A A G A T A T A C A A A G C A T G T A T T A C A A 3936
187  h r v a l t y r v a l t h r h i s a s p g l i n g l u a l a m e t a l a i l e s e r a s p g l n 203
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3937  C T G T T A T G T T A C A C A T G A T C A G A A G A C A C T G C A G A T T A T G A T A G A 3986
204  I l e a l a v a l m e t l y s a s p g l y v a l l e g l i n g l i n l e g l y a r p r o l y s g l 220
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
3987  A T A C C T A T A T G A A A A T G A G A T T A T A C A C A G G T T G A A G C C C T G T A A G 4036
220  u l e u t y r h i s l y s p r o a l a a s n g l u p h e v a l a l a t h p h e i l e g l y a r g t 237
:::||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
4037  T A T T A T A C A A G A C C T T A A T A T G A T T A T T G T C A C T T A T T A G A C A T T 4086
237  h r a s n i l l e i p r o a l a s n l e u g l u l y s a r g s e r a s p g l y a l a t y l l e 253
:::||||:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
4087  C A A T T T A T T C T A T G C A C A T A C A A A T A G A A G A A A T G A T A C T A C T T 4136
254  v a l p h e s e r a s p g l y t y r a l a l e u a r g m e t p r o . . . A l a l e u a s p g l n v a 269
:::||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
4137  T T A T T A G A T G C G A T A T A A A T T G A A A T G A T A A T T A T T A T A G A T G T T A A 4186
269  I g l u g l i n l a l l e h i s v a l s e r i l e a r g p r o g l u g l u p h e i l e l y s a 286

```



```

4187 AGACGAGATGAAGTGTGTTGGAATAAGCCTGAAGAATTTTGTCTA 4236
286 spLuserGlyasp.....IleGlugLythrIleargaspSerValTyr 300
4237 ATTCTGAAATGATGAGAAAGATCAAGCAAAATCTTATCCAAACCTTTC 4286
301 LeuGlyLeuAsnThrAspTyrPheIleGlu.....Th 311
4287 CTTCGAAATATATCTATATTTCTTACATTCATGATATGACGTGT 4336
311 rGlyPheAlaSerLysIleGlnValSerGluSerThrPheGluGluA 328
4337 TCCGAGTCAGCCCTAGCATGAAATATTCACAGACATCATCACAGATA 4386
328 spLeu...GlnYsgLysAsnArgIleargLeuArgIleAsnThrGlnLys 343
4387 GAATGATGAGAGAAAGATGAAATTAACCTTAAACCTAATGCTAATAA 4436
344 LeuAsnIlePheSerAlaAspGlySerGlnAsnLeuIleLysGlyVal 359
4437 ATTAATGATATTACTCCAGATATGAAAAAGTCTTATTAAAGAGATA 4484

```

seq\_name: gb\_ba:AE009386

seq\_documentation\_block:

LOCUS AE009386 10356 bp DNA linear BCT 20-DEC-2001  
 DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 156  
 OF 187 of the complete sequence.

ACCESSION AE009386 AE008689

VERSION AE009386.1 GI:17743066

KEYWORDS

SOURCE

ORGANISM

Agrobacterium tumefaciens str. C58 (U. Washington).  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

REFERENCE

AUTHORS

1 (bases 1 to 10356)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,

Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Boyee Sr., D., Chapman, P., Clendinning, J., Deatherage, G.,

Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,

McClelland, E., Palmeri, A., Raymond, C., Rouse, G.,

Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,

Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,

Krispan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,

Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.,

and Nester, E.W.

The genome of the Natural Genetic Engineer Agrobacterium

tumefaciens C58

Science 294 (5550), 2317-2323 (2001)

11743193

2 (bases 1 to 10356)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,

Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Boyee Sr., D., Chapman, P., Clendinning, J., Deatherage, G.,

Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,

McClelland, E., Palmeri, A., Raymond, C., Rouse, G.,

Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,

Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,

Krispan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,

Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.,

and Nester, E.W.

Direct Submission

Submitted (27-SEP-2001) Department of Microbiology, University of

Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA

98195-7242, USA

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

gene

```

/gene="Atu4577"
173..1165
/gene="Atu4577"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate binding protein"
/protein_id="AA145371.1"
/db_xref="GI:17743067"
/translation="MKSCITLFAFLVAGATPLASAFNAHAGSYVYVSAPDQIMDEL
LPFEKKTGTVEVLVAGSGELMNRKAKAGDVYMSVDGYIDFSADEPYK
VEASINPAFSPSTWVPTAVTAFLVYKRALGAPVPTSMADLAKPEYGLISAR
ADSGSAIOMATVYIDDDSEKGEIYTGINSVLTSSGAADPRVNDGEOAVGIT
LEDAIRYKLGSAPEIYVPEKGTAIADGMAIVGPGNSGKAFIDFYKRAOEY
VYKARKSVRTDVPANVALVPLSDVPDAKYPKRNADNRSLMEKNDIILDVO"
1265..2362
/gene="Atu4578"
1265..2362
/gene="Atu4578"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein"
/protein_id="AA145372.1"
/db_xref="GI:17743068"
/translation="MEOHMAAIVETTSIKSRDYVALSDINISIPSGFTLLPSPGC
GKTLTLRTIAGEHODSGSISIERQAIETHVANHRRDVGMPQDAVAPHISVFNIAF
GLKORSSSAETREVRGKILDVOLAPAKMPHELSGOQORGLAVALYINPKVL
MDEPLNSDAKTRVLDRELREIOAMITVYVYTHOEEALMSDLCWAGYVIOO
AAPREYVNNANRNVASFVANNFLTRDRTGGOTSISLAVTLTPQDAISKDRITVA
AIRPAISIGPSIGNCDEKIDLPYTRGVSTGRMNVAAVLSGEETAITKPSPEI
IALQPNOKTFPSFAGDQTLFGDVGTERLS"
2359..4044
/gene="Atu4579"
2359..4044
/gene="Atu4579"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, membrane spanning protein"
/protein_id="AA145373.1"
/db_xref="GI:17743069"
/translation="WTDMAATRTGGERLRLPGEWTSYVFAVAILLAILPLIFSVF
FISFEDAKTGAFGFSNVAEFTFRRYVALNNTVLIGVLMGACLGIPLAFCTSRY
RIKGTFLATFAVLVLCAPFIGAAYAMLEGCANGAVNLSFAGISLPTIYGIV
WFSLSKEPPIYIMTENDALNTINKSYEDAAVNLGCTAFORRKTILPLVPAVSGAI
ICFVLSIDRCGPATLKGITLSTIAAOCYTSBAGCPMAVITSMVIAISMAAIL
LORHLARRYASSLITNRPVKOSMPLSFLVHGCTYLVITAMLPSTLVTYTSFLAT
SGPVFTGGFGLDSYSRLRDSPOALNSFMFLAAVLAIVSGLSLTVVRDWAAS
GSLDILLWAPYLIPGVMAICYVTFRPPDIDGTALIVLIVLIRLPYGVASTTS
ILROKPIPEEAVALVGLASPAKFTLSVPLMLPGLIIGLSMSTFAINESTSLIY
TARTITMPLVLIQVIDEFGTAALSLILLVSVGCVFVFRLSEDKAEAFV"
4126..5082
/gene="Atu4580"
4126..5082
/gene="Atu4580"
/note="putative; ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA145374.1"
/db_xref="GI:17743070"
/translation="MNIAFSOSKRYKNAATPGDDVAILLPGRVAFPDGATPGANV
DGLSSGRIALAAANAVAMSWISKGTLCNCESTLFTQITSDILKKKEAENSHPSST
LAIVADKMSLRILLAGSGIRINDRLLOSKLIDRSTARSARLHVHQLQASGDD
AAEAKTRATITSGLRAAVKAIGIMKGVDEKYLITRAASGLQNELEAEEFLDMGIVY
QPHYANRGHVLYGASINGVMGEVGLDVLADTVNSLDFIDGYLSTLPTGIDVVD
WEAEFARVETADPHKLNKYPVSKGTSSEFDDRVVCMKAR"
complement(5273..5803)

```

gene  
CDS  
/gene="Atu4581"  
complement(5273. .5803)  
/gene="Atu4581"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA145376.1"  
/db\_xref="GI:17743071"  
/translation="MVNHAHILPLKTSPTISPREKDAAGLAFTVAATVASHAGG  
IGRTITSMPLISAEPRLMISTIDASSRVLDIASRRPVAALSRGOELADYAPGGR  
NLPDRFANGSDWVPSSGSPRLAALLSMDCELVSDIAADHILFVGHILATCDPLRE  
VLMSEKNTKGAEAR"  
complement(5848. .6183)  
/gene="Atu4582"  
complement(5848. .6183)  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ferredoxin"  
/protein\_id="AA145376.1"  
/db\_xref="GI:17743072"  
/translation="MPYVTENCIACKYMDCEVQPEYCEYEGENMLVIHPDCCIDCG  
ICERSCPAALIRPDEAGLHVMDLNRHYSIGWPRHQKRPDNDTMNGAAKFSI  
LSKNPGAGS"  
complement(6288. .6887)  
/gene="sodB"  
/note="Atu4583"  
complement(6288. .6887)  
/gene="sodB"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="superoxide dismutase [Fe]"  
/protein\_id="AA145377.1"  
/db\_xref="GI:17743073"  
/translation="MAFQPLDPLPYTHNALASGMSFEETLKPHHDLHKHKAAYVDNIKNKI  
TGTENKSLSDIYRGITYEKGAVAQSGTFNNASQHRHNNLFWELMGQNIAITSPLED  
ALTSQFGLVAFKQNFALAGASQFSGGAWLVPTDGLKIKITTEGCVNPLCGQAL  
LGGDWEHSHYVLEETKRPAYENFLDNLVMOAVARL"  
complement(7052. .7621)  
/gene="Atu4584"  
complement(7052. .7621)  
/note="Atu4584"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, TetR family"  
/protein\_id="AA145378.1"  
/db\_xref="GI:17743074"  
/translation="MPRRKSIDQVLDALLPLMLQTPDGLFFAAAKAGLSAATL  
VQRYKRDLDVETVILRAMDRLRAETETADEEPLVTGQALISLLRLMPOSAEQNAS  
DGLLRDLIRNPKLRAGALMGHTLHAALGRRLASPKDRERLRGMQMAVWGCAHTW  
MAFVGEAGASIRALAEWAVVNVDER"  
7668. .8195  
/gene="Atu4585"  
7668. .8195  
/gene="Atu4585"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA145379.1"  
/db\_xref="GI:17743075"  
/translation="MAVIRGYIASLDOQIVSGSGSDMLFKKDDMDLGHDRPLFK  
GIRTVKGRATYDFIASEQSPWVYGDQVQIVVIRSPLEERGPLEVIVNDGSLIREIR  
ALDDVDVMLGGGKIQMAFLERGALDEIEIETVPEILIGSEGPLFPLTGHASPTLVSA

```

TAIEKGCYRLHYRBS"
gene      complement(8281..9348)
          /gene="Atu4586"
          /complement(8281..9348)
          /gene="Atu4586"
          /note="identified by sequence similarity; putative; ORF
          located using Blastx/Glimmer"
CDS
alignment_scores:
  Quality: 731.50      Length: 382
  Ratio: 2.729        Gaps: 8
  Percent Similarity: 70.157  Percent Identity: 42.408
alignment_block:
  US-09-769-162 x AE009386 ..
Align seg 1/1 to: AE009386 from: 1 to: 10356

1 MetSerGluLeuLysIleIleAsnAlaLysLysIleThrIleAspValPr 17
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1277 ATGGCCCGCCCGTAGAAATACAGCATCAACAAAGATTACCAGCGGTGCGT 1326
17 ovalIeGluAsnLeuAsnIleThrIleProlGlySerLeuPheThrL 34
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1327 TGCCTCTCCGACATCAATATCTCATATTCATCAGATGCTTTTTCACGC 1376
34 euIeugIAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1377 TTTCTGGACCCAGCGGTTGTGGCAAGACACAGCTTGTGGCAGCATGGCC 1426
51 GlyPheAsnSerIleGluGlyGlyLeuPheThrPheAspAspThrLysI 67
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1427 GGTTCACACGACGAGCAGCGGAGCATGACATCGAAAGACAGCGCAT 1476
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1477 AGACACAGTGCCTCCCAACAGCCGATGTCGGCATGCTGTTTCACAGATT 1526
84 yAlaIlePheProIleLeuThrValAlaArgAspAsnValAlaPheGlyLeu 100
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1527 ATGCAGTCTTCCGCATATTTCCGTAATTCGACAAACATTTGGTTTCGGTCTG 1576
101 MetGlnLysLysValProlGlyGluLeuIleGlnIleThrAsnLysTy 117
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1577 AACACGCGCAAGTCATCTTCAGCCGCAATCGGCAACGGGTGGCAAGAT 1626
117 rLeuGluLeuMetGlnIleAlaGlnIleAlaAspArgLysProlAspLys 134
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1627 TCTTGAGTGTGTCAGCGTGTCTCCCTATGCGCAACCGATCCCGACCAAC 1676
134 euSerGlyGlnGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaVal 150
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1677 TCTCGCGCGGCGACAGCAGCGTGTGTGTACTAGCGCGCAGACTGTGTATC 1726
151 AsnProSerValLeuLeuMetAspGluProlLeuSerAsnLeuGlnAla 167
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1727 AACCCCAAGCTCGTTGATGAGCAAGCAACCTGTCCAACTGTGACGCCAA 1776
167 sIeuAlaIleAspMetArgGlnAlaIleArgGlnIleGlnHisGluValG 184
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1777 ACTTTCGGTGGATTGGCGGCGGAACGTGCGGAATCCACAGAGGCCATGA 1826
184 LyIleThrThrValThrValThrHisAspGlnGlnAlaIleAlaIle 200
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1827 ACATCACACACCGTCTATGTCAAGCATATACAGAAAGGGCGTTGCCATG 1876
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnIleGlyArg 217
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1877 TCCGATCTGGTTGTCGTCATGTATGCGGCGGTCAATCAGCAGGCTGCACC 1926
217 gProlGlyGluLeuThrIleLysProAlaAsnGluPheValAlaThrPhe 234
||||| :|||: ||||| :||| :||| ||||| ||||| |||||
1927 GCCCTGGGAAGTCTATTAACAATCCCGCAACCGCTTTGGTGGCTCCCTGG 1976

```

```

234  legYargThrAsnIlelleProAlasnlenglulysArgSerAspGly 250
      ::||| ::|||::: ::| ::| ::| ::| ::| ::| ::|
1977  TGGAGGACCAATATTCCTCAGCGTCGAC.....CCGACGGGTGGG 2017
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
251  AlatYrIleValPheSerAspGlyTyrAlaIleuArgMetProAlaLeuAs 267
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2018  CAACATCATCATTCACG.....CTCAGGTGACGCTGCCGCGAGCGGA 2061
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
267  pGlnVal...GlulGlnIleAlaIleHisValSerIleIleArgProGlu 283
      | ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2062  TGGGATTTCACAGGACCGAGCATGCTGCTGCATCGCCCGAGAGCGGA 2111
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
283  llele.....LysAspGluSerGlyAspIleGlu 292
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2112  TTTCTATGTGCTCCTCCATAGGCAATTGTATGATGAGCCATCGAATCCCC 2161
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
293  GlyThrIleArgAspSerValTyrLeuGly.....LeuAsnThrAspTy 307
      ||||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2162  GTGACCATCAGCGACGTCAGCTTCACTGCTGTCGATGAAACGTGCGCGC 2211
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
307  rPheIleGluThrGlyPheAlaSerLys..... 316
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2212  CGTCTCTCTCGTGGGTAGGAAATCGAGCGCATCCAGACCATCGCCGG 2261
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
317  .....IleGlnValSerGluSerThrPheGlu..... 326
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2262  AGATCATCGCGCTTCAGCCCAACCAAGAAACGACGTTCTCTCGTTGGC 2311
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
327  .....GluAspLeuGlnLysGlyAsnArgIle..... 335
      ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2312  GCGCATACGCAATGTTCGGGTGAGGTGTAAACGGGAGCGGCGTGCATG 2361
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
336  .ArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
      ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2362  ACGGATATGCTGCGACTGCACGACGCGGGGCTTCCGCTCGCAGC 2407
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
seq_name: gb_ba:AE008229
seq_documentation block:
LOCUS AE008229 11697 bp DNA linear BCT 18-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 33 of
187 of the complete sequence.
ACCESSION AE008229 AE007870
VERSION AE008229.1 GI:15158623
KEYWORDS
SOURCE Agrobacterium tumefaciens str. C58 (Cereon).
ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 11697)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11697)
AUTHORS Hinkle,G., Slater,S.C. and Goodner,B.
TITLE Direct Submision
JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
Street, Cambridge, MA 02139, USA
COMMENT Approximately 800 bp of telomeric sequence missing from the left
end of the chromosome and 200 bp missing from the right end.
FEATURES
source
1..11697
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/db_xref="taxon:181661"
/complement(70..900)
/gene="AGR_L_572"
/complement(70..900)
/gene="AGR_L_572"
/note="hypothetical protein"

```

```

/codon_start=1
/transl_table=11
/product="AGR_L_572p"
/protein_id="AAK88858.1"
/db_xref="GI:15158624"
/transl_table="MLKFSAPFAVAWAFSATPAAGADQVAATPAGVDGPDMAFOATGYL
WATGNGINISPEFORAGTLHVEKFSFSDVDDLDFGFLNIGWGYDFEYSGDMYVDTT
DSKAGPLPALOIPLPTLPACATIDANVDQEFMATLOAGYRIIDADGFSLDALAGA
RPMHTSNDVAVYASHSLIGSVATHRRESQPDVYIGARAPESLTDXLSVGGQVNDIG
FCAGSDLTMSALATAYNYIFTDTLSAGIKYLDVYSHDGYFDSRLSGPVGLTYRF"
1009..2076
/gene="AGR_L_573"
1009..2076
/gene="AGR_L_573"
/note="probable hydrolase YCA0877 [imported] - Vibrio
cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_L_573p"
/protein_id="AAK88859.1"
/db_xref="GI:15158625"
/transl_table="MRPKTLPEFSACTPAIESIAFLTTSVAEACTREYVIGENNOVMT
ARSDMKTDVGTNLWVFPKMERSEAGPNSVKMTSKYGSYIASGYDVSTDGNNBAG
LAANYLMIVESYYPDYDGKSPGLSIAAMQYLDNFATVEEAVYLEKNPFIIVDSV
PGEERLATLHLISDASGDSATVEYIDGQVTHHGRQOVMTNSPTPEOALANWT
OIGCTVM.LPGRNASDRYRASFYANAIIPKSNPEALIASVPSYRMTSVPGITTPD
QPNISTRRRYVIDHKKLIYFESALTPVFWIDMTKIDLKETAQVKKDLGANOIH
IYSGANESLKDTKPKFPLGL"
/complement(2162..2689)
/gene="AGR_L_575"
/complement(2162..2689)
/gene="AGR_L_575"
/note="conserved hypothetical protein PA3469 [imported] -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_575p"
/protein_id="AAK88860.1"
/db_xref="GI:15158626"
/transl_table="MAVIRGYTAASLDQIYSGDGLMFKRYDDMDLGEHDYRLF
GIRTVMGATYDFIASQSPWVGDQRYIVTSPRIEPPKGLPVYNDVGLIELR
ALDDGDVMMGLGGKLOMAFLERGALEIYIYIPILSGEPLFLTGHRASPTLVSA
TAIEKGVRLHFRFS"
2736..3305
/gene="AGR_L_578"
2736..3305
/gene="AGR_L_578"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_578p"
/protein_id="AAK88861.1"
/db_xref="GI:15158627"
/transl_table="MPRKSISDDOVLDPMLPMTQSGDGLTFAAACAAGLSAATL
VORXGRDIEVETVLIRAMDRLRAETREDEEPELISQASLILRLIMPQSAEONAS
DGLILLRDIRNPKIARAGALMGLTILAHILGRILASDKDRARLGMQAAVVOGAHTW
MAFVRGEDGSAIRRLERVAIVVNDER"
3347..4069
/gene="AGR_L_579"
3347..4069
/gene="AGR_L_579"
/note="(AF022931) superoxide dismutase"
/codon_start=1
/transl_table=11
/product="AGR_L_579p"
/protein_id="AAK88862.1"
/db_xref="GI:15158628"
/transl_table="MDLPKRLRTSTVLEVLVSPCPKALSMILSRQITKEYMAF
QLPDLPTTHDALAPSGMSERTLKPHHDLHKKAYVNDLKKATITGEMKSSLDIYRGT
YKGAVAQSGITFNNASQHNHNLFWELIKPQNHAIPLSPLELDLTSFGVALFKONFA
LAGASQFGSGMALLVVDITDGLKILRTENGVPNLCFGOKALLGCDVWHSYIIDRTK
RAYVENFLDNLVNMQAVARL"

```

```

gene      4162..4509
CDS       /gene="AGR_L_582"
          4162..4509
          /gene="AGR_L_582"
          /note="(AF28062) ferredoxin II"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_582p"
          /protein_id="AK88863.1"
          /db_xref="GI:15158629"
          /translation="MSIPMPYVVTENCIAKCYMDCEVCEPCEYEGENLVTHPPDC
          IDGICERCPAAIRPDTPAGLHVLDLNRHSGIMPVHQKRTPEPDNATMNGAAG
          KESILSKNPGAG"
          4521..5084
          /gene="AGR_L_583"
          4521..5084
          /gene="AGR_L_583"
          /note="(MADH:FMN OXIDOREDUCTASE)"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_583p"
          /protein_id="AK88864.1"
          /db_xref="GI:15158630"
          /translation="MSHNLRLIDGCVNNNAHILPKTSTPTISPREKDAAGLAFTV
          AVATSHAGERIGRTITSEPLSAEPPLMISIDASSRWDLIAARRRSVAALSGQ
          EEIADVFAGKGNLPDRFAVGSMDVWPSGSRFLAALLSMDCELVGSIDADHILFVGA
          IIEATCDPLREVLIMSESRNYKGAASR"
          complement(5275..6231)
          /gene="AGR_L_584"
          complement(5275..6231)
          /gene="AGR_L_584"
          complement(5275..6231)
          /note="(hypothetical protein)"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_584p"
          /protein_id="AK88865.1"
          /db_xref="GI:15158631"
          /translation="MNIASFQSKRYKNAATPGDDVALILRGVFAVFDGADTPTGANY
          DGLSSGRIALAANANVAEMSIKGTLCNCESTLPTOTISDLKEAERINASHPPTT
          LAIVADMGSLRLLLAGDSGIRINGDRLOHSLIDRVSTARSRIHVFKOLFAFGSD
          AAARAKRTTFSGLRRAAVKAGIMKYDAEKVITTAASGLQNEIETAVEEFLMGIRV
          OPIYANRQGHVLTGASINGSMVEESTVDLSAKDVNSLEIFTDGILSLPTGIDVVD
          WEAEFARVETADHKLINEYPSVKGSTSEFSDDRTVYCMKAR"
          complement(6313..7998)
          /gene="AGR_L_586"
          complement(6313..7998)
          /gene="AGR_L_586"
          /note="(586"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_586p"
          /protein_id="AK88866.1"
          /db_xref="GI:15158632"
          /translation="MTDMAATATGGRFLRPGFWTSVVFVAILAIFLPIESVE
          FLSFDPAKTGAEFSNVAEFTRRFTVALMNTLVIGLGMGACILGLPLACTSRY
          RIKGRFIATFALVLCAPFEGIAVAMILFNGANGVFNLSFAGISLPIEIGPVI
          MYSELEPFYIYLTEMENALNTINKSYEDAENLGCYAFRFRKITPLVEPAVSTGAI
          ICFVLISADGFPALIGKIRLTSTIAVOYTEMAGTPTMAVTISMVIAISMAALL
          LORHILAKRYVASITLNPVKOSMRPKISFIYVRCYIVAFETMISLIVIVISPLAT
          SGVPTFGGGLDSYKRILKDSPOALNSFMFALAAYVLAISGLSFLIVRDNVNS
          GSLDLAMPYPLIPGVNVAIGVTRYPPEPDTGTALLIILVLFIRLPIYGVSTTS
          ILQDLIPSEBAVNAGSPAKTFLSVTVPLMPLGLISLMSFITAINELISSTLILY
          TARTIMPVLIVIOVIDEGFGTAALSTLLVSTGCVFVFRSLSDKKEASFV"
          complement(7995..9119)
          /gene="AGR_L_589"
          complement(7995..9119)
          /gene="AGR_L_589"
          /note="(U75349) putative ABC transporter Btld"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_589p"
          /protein_id="AK88867.1"

```

```

          /db_xref="GI:15158633"
          /translation="MVACRAGYMEQHMAVEITSIKSYRDVVALSDINISIPSGSF
          FTLLGSGCGKTLTLRTIAGFHQODSGSISIERQALIEHVPAHKRDGVNFODYAVPPI
          ISVDNIAFGIKOKRSSSAEIRREVAKLIDVVOALYAKRMPHEISGGQOQRYGLARA
          LVINPKVILMDEPISLNDLAKRLVDRRELRIOOANITTVYVTHQOEALAMSLDYC
          VMVGVIQOAAAPPEVYNNRANRFAVASPGANNEFLDRTGQGTSSISIKVLLPOADA
          ISKORTIVAAIRPEALISIGPSIGNCERIELEPPTIYOVSTFGRENNVAVALSSEIE
          ATRKPSPEITAIQPNQKTFPSWFGDTQLFGDGVTERLS"
          complement(9192..10355)
          /gene="AGR_L_591"
          complement(9192..10355)
          /gene="AGR_L_591"
          /note="(U75349) periplasmic-iron-binding protein Bita"
          /codon_start=1
          /transl_table=11
          /product="AGR_L_591p"
          /protein_id="AK88868.1"
          /db_xref="GI:15158634"
          /translation="MFRITVHFNFNEFALDLLIVYKINSRQRAVTTKNSBKLSRG
          TSCVRQKNGAPFMQSCYIFALVAGATFLASAAFAHAGETVYVYSAAPQDLMDLL
          PLFEKKTGKVELIKAGSELNMRIKAEITGAAGDVIMSVDTGVIDFSADLEPYKPV

alignment_scores:
      Quality: 731.50      Length: 382
      Ratio: 2.729      Gaps: 8
Percent Similarity: 70.157 Percent Identity: 42.408

alignment_block:
US-09-769-787-162 x AE008229/rev ..

Align seg 1/1 to reverse of: AE008229 from: 1 to: 11697

1 MetSerGIuileIySillelleasnaIleasIlySilleYrhIsaPvalPr 17
||||| :||||| :||| :|||||
9080 ATGGCGCGCCGAGTAAGAAATTACACGATCAAGAAAGTTACCGGGATGTCGT 9031
17 ovalIleGIuIsaIleuasnIlethIleProIyGlySerIeupherhL 34
||||| :||||| :||||| :||||| :|||||
9030 TGCCTCTCCACATCAATCATTCATTCATCCATCCATCGAGATCGTTTTCACGC 8981
34 euleGIyAlasergIyGlySyrThrIleuArMetIlela 50
||||| :||||| :||||| :||||| :|||||
8980 TTCCTGGACCCAGCGGTTGTGGCAAGCACGCTTGGCAGCGATGCC 8931
51 GlyPheasaserIlegIyGlyIuPhetYrPheaspaSprhIyI 67
||||| :||||| :||||| :||||| :|||||
8930 GGTTCATCCAGACGAGCAGCGGCATCATCGAACAACAGCGCAT 8881
67 easnAsMetGIuProSerIySArgasnIlegIyMetValPheGIuasnt 84
||||| :||||| :||||| :||||| :|||||
8880 AGAGCACGCTGCCCTGCCACAGCGGATGCGCATGCTGTTCAAGATT 8831
84 YrAlaIlePheProIIsleuThrValArgaSPasnaValAlaPheGIyIleu 100
||||| :||||| :||||| :||||| :|||||
8830 ATGCAGTCTTCCGATATTTCCGATTCGACACATTCGCTTGGCTCG 8781
101 MetGIuIySValProIySgluIleuIleGIuInThraSnIyStY 117
||||| :||||| :||||| :||||| :|||||
8780 AAACAGCGCAGATCATCTACCGAGATCCGCAAGCGGCGCAAGAT 8731
117 rIeugIuIeumetGIuIleagIuIyAlaSPargIySProaspIySt 134
||||| :||||| :||||| :||||| :|||||
8730 TCTTGATGTCCTGCGACGCTTCCCTTACCGACGATGCGCAGCAC 8681
134 euserGIyGIyGInGIuInArgValThrIeuaIacyAlaIeuaVal 150
||||| :||||| :||||| :||||| :|||||
8680 TCTCGGCGGCGCAGAGCAGCGTGTGCTAGCGCGGCGCATGCTATC 8631
151 AsnProSerValIleuIeumetAspGIuProIeuserAsnIeugIuaIy 167
||||| :||||| :||||| :||||| :|||||
8630 AACCCCAAGTCCTGTTGATGAGCAACCGCTGTCACACCTTGACGCCAA 8581
167 sleuArgIeuaSPmetArgIuaIaIleArgGIuIleGIuInIsgIuValG 184

```

```

8580 ACTTCGCGGATTTGCGCGGGAAGTGGTGAATCCAGACGACCATGA 8531
384 LyllethrhthrvAlYrvalThrhIsAspGlnGluGluAlaMetAlaIle 200
8530 ACATTCACCAACCGCTATGTCACGATGATCAGGAAGAGCGCTGGCATG 8481
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
8480 TCGGATCTGCTTGGCTCATGATGCGCGGCTATTCACAGAGCTGCACC 8431
217 gProLysGluLeuTyRHisLysProLysAsnGluPhaValAlaThrPheI 234
8430 GCGGTGGGAAGTCTATACATCCGCCAACCGCTTGTTGGCTCCTTGC 8381
234 legIYrgrHrAsnIleIleProAlaAsnGluLysArgSerAspGly 250
8380 TGGGACCAATATATTCCTGACGCTGCAC.....CGCACGGGTGGG 8340
251 AlaTyRlleValPheSerAspGlyTyRAlaLeuArgMetProAlaLeuAs 267
8339 CAACATCCATATCCAGC.....CTCAAGGTACGCTGCCGACGCGGA 8296
267 gAlnVal...GluGluGlnAlaIleHisValSerIleArgProGluGluP 283
8295 TGGATTTCCAGACGACCGGACATCGTGGCTGGCATCCGCCGAGAGCGA 8246
283 heIle.....LysAspGluSerGlyAspIleGlu 292
8245 TTTCTATGTCCTCTCCATAGGCATATTGTGATGCATCAGCAATGCCCC 8196
293 gLyThrIleArgAspSerValTyRleuGly.....LeuAsnThrAspTy 307
8195 GTACCATACGACGACGCTCACTGCTGCTGAGATGACCTGCCGCC 8146
307 rPheIleGluThrGlyPheAlaSerLys..... 316
8145 CGGCTGCTGCGGTGAGAAATCAGGCGCATCACCAACCATCGCGG 8096
317 .....IleGlnValSerGluGluSerThrPheGlu..... 326
8095 AGATCATCGCGCTTCAGCCAAACCAAGAAAGAGCTTCTCTGTTGCG 8046
327 .....GluAspLeuGlnLysGlyAsnArgIle..... 335
8045 GCGGATACCAATTGTCGTTGACGCGTACACCGGGAGCGCTGTCATG 7996
336 .ArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
7995 ACGGATATGCGTCGACATCGACAGCGCGGCTTCCGCTGCGAC 7950
seq_name: gb_ba:AE001779
seq_documentation_block:
LOCUS AE001779 20990 bp DNA linear BCT 02-JUN-1999
DEFINITION Thermotoga maritima section 91 of 136 of the complete genome.
ACCESSION AE001779 AE000512
VERSION AE001779.1 GI:4981777
KEYWORDS
SOURCE
ORGANISM Thermotoga maritima.
Bacteria; Thermotogales; Thermotoga.
REFERENCE
1 (bases 1 to 20990)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
Eisen,J.A., Fraser,C.M. et al.
TITLE Evidence for lateral gene transfer between Archaea and Bacteria
from genome sequence of Thermotoga maritima
JOURNAL Nature 399 (6734), 323-329 (1999)
MEDLINE 99287316

```

```

REFERENCE
2 (bases 1 to 20990)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
TITLE Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..20990
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
complement(97..1203)
/gene="TM1224"
complement(97..1203)
/gene="TM1224"
/note="similar to PID:1208894 SP:044406 percent identity:
54.86; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, xylR-related"
/protein_id="AAD36299.1"
/db_xref="GI:4981778"
/translation="MPKSVRAENISRLIKRMKSPVSVLEAEELGTLKTVGEIARI
FLERGIYVEEDSPKRGKRPKSLKISPNCAVYVIGIEYTRREIACGLIDASMNILAE
AHLPSQSDREETLVMRIITDRAKDMKELGSKLSALTVAAPRIDEREITIDIPRN
FPLSQIPLANLKKRYGLEWVENDADGAVGEKWKYRKDSFAMPIITLKGIGAGIIT
DGEIYRGSENGAGEIGYRVNGNDEYVLEBDVCENNVYLKHLVMSGESLLEADSDG
VRKEYFDDIARFYSIGLNLHLFGISKIYVIGGFELGKNGENFLKIKIEVETHLYK
HSVMSKSPVOEPIYAFGAHALENTYERTTS"
complement(1217..2197)
/gene="TM1225"
complement(1217..2197)
/note="similar to GB:Pyro_h percent identity: 52.69;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAD36300.1"
/db_xref="GI:4981779"
/translation="MKVTEKIPNIPMEERPEGYGPVWRYSKNPIIGRNVPGARV
ENSAVVPNGSEFVFRIDKNTREPLIFEGSKGGINNEIPEEIQWVDVNGEPQPS
YAIIDRVVAKIEDTYITITCDHGEPTIGVGTAKDKTFVRLPNATVPNRNGVLPK
INGKVMNLRPSDNGHPEGDIFLSESDMTLHWGNHREVLGRSSYNMWNELKIDAGPY
PIETSEGMLLIYHGVTLLNGVYVSGAALIDLDDPSKVLRLRSRYLLTTPREEYETVG
FVPNVFPCALCDADTGRVAIYVGAADTHVALAFGYIDEIVDFKRRSM"
complement(2250..3938)
/gene="TM1226"
complement(2250..3938)
/note="similar to GB:Pyro_h percent identity: 57.09;
identified by sequence similarity; putative"
/transl_table=1
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein, putative"
/protein_id="AAD36301.1"
/db_xref="GI:4981780"
/translation="MNMRRLILPVYVFLFSSLSQVLERNETWYGGSLSPSSNNN
PPTRNNAVPGTGLVETMFFYDPLTGFDPMLAEKGMELDSKTYRVLRLGTYWHDN
VPLTSEDVRFTELAKRYKGIHYSVWMLDIETPDNRVILVFKDRIYHEMELLT
TLPLPKHIMEKETTLLOSNEVPLSGSVVAHSWQNMKIFERFENMWGTVMYG
KPARVYVILVELSNVALGMLKGEIDFSNMLPGVILKKVNLNLTXYDEPYNHLS
SYVGLFNARKYPLSLPEFRRAIMSNADPIYORVYEGAVLADPLGFPYSSVMK
YKPEVVEKHGEKYPDEPAKSLIDKLGFRDVGNGGFRFPDGKPKILTECPGWTQY
MOAIVYIDQIAKVGINAEYFPDSSKTYEMNKYGEPIEMNANGTGSSTPYRYNT
IIFPDALSEFSYTGNTGRVONPEVSELEELNRPIDLVNRYELGCLGELIILKDL
FPIPLWGAFAFIQDNYWTWPNENHNPVAMPCCWAMWQIGALKILFENLKPAA"
complement(3935..5944)
/gene="TM1227"
gene

```

CDS complement(3935..5944)  
/gene="TM1227"  
/note="similar to PID:1679597 percent identity: 54.44;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="endo-1,4-beta-mannosidase"  
/protein\_id="AAD36302.1"  
/db\_xref="GI:4981781"  
/translation="MRFMEILLIVELSVLPASDEFKAVENGKALNCKEERETISN  
NYMKSNRMIDSVLESARMDGKIKLVIVNNMDDGGMNQVVRNGGTHHDFYDE  
ISNAOGEPRLDIYAKAKELGICLVIVNNMDDGGMNQVVRNGGTHHDFYDE  
KIKREYKATVSVLNVHNTYTGPTREEPTIMAMELANEPCEDTSKNTLYEAKEM  
SSYIKSLDPNHLVAAGDEGFNSNIEGFKPYGGAEWAMVSGVDMKLLSTLYDQEG  
TFLHPSHWPBENYAWQAWKMEIDHIAKIKGKPVVLEEGIIPKAPVNTALYR  
LNNDLYVDLGGGAMFWMLAGIGESGDRDERGYPPYDGFRIYNDSPAEELIREAK  
LNTGEDIREDTCSFLLPKDGMIEKKTIVRAGVDFYSMTPEKLSKVDLYEENIE  
HIGYGTGFDLDTTTRIPDGEHEMFLLEGHOCQYKDSIKAKYVNEARVYLAHEVPS  
PEVKRMNNSGTWQAEFGSPDIEKNGEVNGALQLNKLPGKSDMEEVVARKFERLS  
ECEILEYDIYIPNVEGLKGRLPYALNPGWVIGLMMNNAVESAILTTGKEYRR  
FHVRIEFDRTAGVKELHIGVVDHLRYDGPFIIDNVRLYKRTGM"  
complement(6103..6903)  
/gene="TM1228"  
complement(6103..6903)  
/note="similar to GB:M81878 SP:E26833 PID:144859 percent  
identity: 55.56; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, RpiR family"  
/protein\_id="AAD36303.1"  
/db\_xref="GI:4981782"  
/translation="MKIRDKILNVYTFQSPARKVADYVLERPDIYHSITEFAKIV  
GSEETIHMRIKKLDEEGYQAKIALARELSGLEETIERDRPIDEIDILRLKIDLD  
MKNVKAVMIISARHILFEFGVLSGVSEVYSIKFSLGFRFTSPNDHYVIAEVN  
LTGEDLVISHTGNIRDTVKSTQVAKDGAKTATTTKOSELAAVHVLVQSPVK  
IDTYEFLRENIEIAVVDYLFETFORIYREKKHFNLEGVFKPRK"  
complement(6887..8335)  
/gene="TM1229"  
complement(6887..8335)  
/note="similar to GB:Pyro\_h percent identity: 69.66;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36304.1"  
/db\_xref="GI:4981783"  
/translation="WESASGSDRSOKARGAFRVNSLQIRYTSRPEIRKOSRMLQR  
ERRCKGREDPEKGRGNFSPLRVSLODGRTOIQRSKELQFRETNRHNEOGIT  
MKVVGIPSYNNAETISHVARTAAOIGVDFGDSGADGTERHMETDTE  
GLPKESFYEGGLGKSGAMRAIMEFALAKODAAVYVLDLDSVPMWVERIAGLPK  
GEADYVTPYLRHREFDGTITNNVCPMPATVLAGKRVKROIGDGFGRRLIYIGKP  
KRIWMDVARFGIDIMTTTAINESGRVQALGCRVHVHQRKQKMGKMEFOVGTG  
PELVITYTENWKEIKIEDVPIYGETPDEPVSMSIDIGNLKLARETILEEVEYIDRG  
LISEVKESGTLSSWVDTLYRSAAVOYRTRTKRVENLLPYFARTAFABEVSLS  
DEEARVYEQDVFLEKHSLEEMKVEDKR"  
complement(8026..9309)  
/gene="TM1230"  
complement(8026..9309)  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36305.1"  
/db\_xref="GI:4981784"  
/translation="MKIALIHRGGLMDVSLSEMKKKVLTIRMGHEVHIVAKNKEG  
VDLTLEIGFENPDERVNRNFGGIDKFLSEKEDFLEKEEELFHTLINDALDYP  
LIVPNINWISIGLPSGLAHSLEKNFVAHHNDKPFMERHNLIPKRRPPELTIDKRP  
DLPNKHVYINTIAQBELKRRRINDSVYVNPVNDSPITSEMIHRVREELQIAPGT

IVALQATRIDRRKTTIELSDVSLKEFTLTKKREADLYNGERYSEVILLSESCIDE  
EYLKELKEVASSGVSLVISEVRKNSLFEKLIADFPVYPSILBDMGQMLEAI  
AAKKPVVLEFEYKFSKDIKPAKLYKNSLDRCFKRVNGLVKVDERILKRAVEISRLF  
DPSLRVEVHEHFEVGRHFSLEERLEDLSREVLD"  
complement(9306..11804)  
/gene="TM1231"  
complement(9306..11804)  
/note="similar to SP:p21139 PID:205299 percent identity:  
49.50; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="alpha-mannosidase-related protein"  
/protein\_id="AAD36306.1"  
/db\_xref="GI:4981785"  
/translation="MKVYVYVNNHMDREMTSSEYTSKMLKEVFPVRELVOKNPEV  
YLDQTADELIVLHPPLEEDLIELVRSGLLGAPYIOIDMHPGASTLKNEFI  
GKDNRRGRBNAMGLDLSRCHISOEOLHPTIEVFLMRGISFENDIGSOEFP  
KSDGTAVOGVTVLGGYRNLVLAETODIAERLKHVEYKLAKEPFRSGEILLDGYDI  
DISPDADYLVNVEIVSPEEPERPENAPTLTLAHNDSGCVADPVEHMERVRYN  
SDLVKRLKMDLSVLNGEELSDFWMDYLTSLAHNDSGCVADPVEHMERVRYN  
YLVKSVESRLHLVERLGLKGTALSLSPSYNHHGCDERTFLKTNGAGLEKT  
EVADEKKEVKELTMENERYRAYFEGDGLVLRKAGVYKLEKEDGAYVTSVEETD  
SLEKMPVLOTKSNVYELERTIRSSOCLITEMREYVTPDETPIYKRIALIKIGKN  
YLSWYETGSGEITRAKMPDYVYKREKDSYLLPPTPEBELKITLAAREGEVREP  
FOGFVSVSEESSKITLARGLKRYMEERKAVNTLSRVENYIAKAEVKRVDAGPMY  
VFGACGEVFTVDLAFMELPVHPRSEEEFRWLLDNDPDLADVADSTEKETFFS  
SLPMAGVQDRYHLAVYNYLIEIEGLRPLQIGKSIESIASEOCVSRNLINPEEP  
VYSVKNPGEVIEFLKETSLEDEITNLBECKDQSRKHLRFLSLRTRKEIELSLQY  
LIDVKTARELNELMKRRTYVYVLEPSEERE"  
complement(11801..12868)  
/gene="TM1232"  
complement(11801..12868)  
/note="similar to GB:Pyro\_h percent identity: 78.03;  
alignment\_scores:  
quality: 716.50 length: 363  
Ratio: 2.559 Gaps: 6  
Percent Similarity: 77.135 Percent Identity: 40.220  
alignment\_block:  
US-09-769-787-162 x AE001779/rev ..  
Align seq 1/1 to reverse of: AE001779 from: 1 to: 20990  
1 MetserGluileuYsileleasAlaLyslySletyrHIsapValPr 17  
12868 TTGGCCGACGTGAATAATGACGGAGTGAAGAAATATTTCCGGAAATGCGAG 12819  
17 ovalileGluasnuLeuAnilethrileProlysglySerleuphetrL 34  
12818 GGCTCTGATGAGATCGACCTGTGCGTGAACGAGGAGAGTTCCTGCTGC 12769  
34 euleuGlyAlaserglylysthrileuLeuAymetlealea 50  
12768 TTTCTGGTCCATCTGTGTTGGAAAGACACCTTCGAGGTGTAAGCG 12719  
51 GlyPheasnserileGluGlygluPheayrPheaspsrthrlsII 67  
12718 GGTCCTGACACGACGACGGGGGAAAGATCTTTTCAACAGATCGCGACGT 12669  
67 easnAsmetCluProserLysarGasnileGlymetValPheGlnAsnt 84  
12668 GACCAACCTTCCTCCAAAGACAGACGACATTCGATGCTCTCCAGAGCT 12619  
84 yralaIlePheProHisleuThyValArgAspAsnValAlaPheGlyLeu 100  
12618 ACCGGGTGTGGCCCATGATGATGATGATGATGATGATGATGATGATGATG 12569  
101 MetClnLysLysValProlysgLluLeuileGlnInthrasnLys 117  
12568 AACTGAAAGAGTCCGAAAGAGAGATGAAAGAGGAGGTAAGTGGC 12519

117 rleugluuemetGlnIleAlaGlnTyraIaAspArgLysProAspLysL 134  
12518 AGGGATCTCCCTTCACATCTCCGAACCTCTCGACAGTACCCGACAC 12469  
134 euSerGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150  
12468 TCTCCGAGGTAGAGACAGAGGTTGCTCCGACGAGCATGCTTAC 12419  
151 AspProSerValLeuMetAspGluProLeuSerAsnLeuGluAlaVal 167  
12418 GAACCCGAAGTACTCTGTATGAGCAGCCCTCTTCCAACTGTGAGCAGCT 12369  
167 sleuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184  
12368 TTTGAGAGGTGAGATGAGAGACGAGCTGAAAAGCTTCAGAGAAAGATTG 12319  
184 lylIleThrThrValTyraIThrHisAspGlnGluGluAlaIle 200  
12318 GAGTCACACCATCTATGTGACTCATGACACAGACAGGCGCATGACGATG 12269  
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217  
12268 GGGACAGAGATCGGGGTGATGAAACGAGGAAAGCTTCAGAGGTTGGAAC 12219  
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234  
12218 GCCCTCTGAGATCTATCACCATCCGGTGAACATCTTCGTCGCTGTTTG 12169  
234 leGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250  
12168 TGGGATCACCGCAGATGAACCTCTTAGAGATGAGATGAGATGGAAGA 12119  
251 AlaTyIleValIlePheSerAspGlyTyraIleAlaLeuArgMetProAlaLeuAs 267  
12118 AATAGCGTTGTCTCCAAAACGGT...GAGATTAAGATTCCCGCGAAGAC 12072  
267 pGlnValGluGluGlnAlaIleHisValSerIleArgProGluGluPheI 284  
12071 CGATCCAGGTGCAAAAGATGATCTTGGTATTCACCCGAGAACGCTCT 12022  
284 leLysAspGluSerGlyAsp.....IleGluGlyThrIleArgAsp 297  
12021 ATCTGAGGAGAAAGCCGAACACTTGAAGCTGAGAAGA..... 11984  
298 SerValTyLeuGly.....LeuAsnThrAspTyrPheIleGluThrG1 312  
11983 GAGGTTTATTTCCCGAAGAACTCATGTCGATGATCTCTCATCTCA 11934  
312 yPheAlaSerLysIleGlnValSerGluGluSerThrPheGluGluAspL 329  
11933 CGTTGGAAGTGAAG...AAGATCTCCCAAAAATTCAGAGGGAATGTGATTT 11887  
329 euGlnLysGlyAsnArgIleArgLeuArgIleAsnThrGlnLysLeuAsn 345  
11886 TCAGAGCGGTGAGAGATCAGCTTTTCTCGACGTTGAAAAGATCCAT 11837  
346 llePheSerAla..AspGlySerGlnAsnLeuIleLys 357  
11836 CTCTTTCACCCCTGAGACGGAGAGAAATCTCATGAAA 11799  
seq\_name: gb\_ba:AF003189  
seq\_documentation\_block: 298050 bp DNA linear BCT 17-JAN-2002  
LOCUS AP003189  
DEFINITION Clostridium perfringens DNA, complete genome, section 5/10.  
ACCESSION AP003189 BA000016  
VERSION AP003189.2 GI:18146727  
KEYWORDS  
SOURCE Clostridium perfringens (strain:13) DNA.  
ORGANISM Clostridium perfringens  
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
Clostridium.

REFERENCE 1 (sites)  
AUTHORS Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,  
Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.  
TITLE Complete genome sequence of Clostridium perfringens, an anaerobic  
flesh-eater  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)  
PUBMED 11792842  
REFERENCE 2 (bases 1 to 298050)  
AUTHORS Shimizu,T.  
TITLE Direct Submission  
JOURNAL Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical  
Sciences, University of Tsukuba, Department of Microbiology; 1-1-1  
Tennohda, Tsukuba, Ibaraki 305-8575, Japan  
(E-mail: tshimizuend.tsukuba.ac.jp, Tel:81-298-53-3354,  
Fax:81-298-53-3354)  
COMMENT On Jan 14, 2002 this sequence version replaced gi:18144662.  
FEATURES  
source  
location/Qualifiers  
1..298050  
/organism="Clostridium perfringens"  
/strain="13"  
/db\_xref="taxon:1502"  
/note="anaerobic pathogen for gas gangrene"  
complement(34..1036)  
/gene="CPE1002"  
complement(34..1056)  
/gene="CPE1002"  
/note="340 aa, no significant homology."  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB80708.1"  
/db\_xref="GI:18144663"  
/translation="MSNKKIRFKMNSNDKVLINSLVLDGSLAKELSKKIDIOKRD  
IIRKNEBGFEDSDGIFPKDKESISIDIKPVEYSKSLPRTPEREDVQSIIDSI  
NEKLNIDIEFTNILLPEIKTSEKVSATKEDNLTQSMTLEIPVSRARRK  
KKKRTLIRMIISYFNKNSKIFLIRKKNKSSKIECDNENKKEVMIKSTNSLDE  
VAIEKLVDNKIEFTSHGINSKLEPEIKIPKFKGRRRIPKVDCLDEINFLNSD  
YSLDKKDFDLKLSDFIKNINSDSLESRRKSLSEBRVCLNLEKIKNLEEFLEIF  
QAMSNK"  
1361..1540  
/gene="CPE1003"  
1361..1540  
/gene="CPE1003"  
/note="59 aa, similar to gnu:AP001515\_87 BH2353 gene  
product from Bacillus halodurans (79 aa); 46.7% identity  
in 45 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAB80709.1"  
/db\_xref="GI:18144664"  
/translation="WNIDELTKRINELHKHKHEGLSDEHKEBELRKEYINRFKSN  
LRQLKGIPEPNKKN"  
1552..2475  
/gene="hprK"  
1552..2475  
/gene="hprK"  
/note="307 aa, similar to gnu:AP001519\_105 HPr (Ser/Thr)  
protein kinase/phosphatase from Bacillus halodurans (310  
aa); 45% identity in 282 aa overlap  
CPE1004"  
/codon\_start=1  
/transl\_table=11  
/product="probable HPr Ser/Thr protein kinase/phosphatase"  
/protein\_id="BAB80710.1"  
/db\_xref="GI:18144665"  
/translation="MGVTIEKLIDFSLLEVIOIGENVPINVDNBPGLQAGFYNY  
FAPRIQVIGAKMSFLEDMSPDLRRKRINFPFSDISCLITGELIHEBELAKRK  
RNWILRSDWYTKRISKITMYLSDKMAPERLRGLVDYVGIGIMLGESGIGKSTP  
ALEIKRGRHLVTDADVDIKETDIDLGRSEITRFGMLPYVAGKMTIDYSALYGSSTL  
NSKQIKITIEHMKDGDYDLGVNDYODILGVKKRLKLVPIRPNRLAVITTEAA  
ANYRYRMSDISPDIIERMLESMESKSI"  
2520..3449  
gene





```

|||||
284497 TGGATATTTTCCCTACCTAAGTAGAAGAAATGACCTTTGGATTAA 284448
101 eGlnIlyslsValProlyslsGluIleuIleGlnIlnhrslsnylTYr 117
284447 AATGAAAAAGAGACCAAGAAAGAAATAGAAAAAGGTTGATGATATA 284398
118 LeuGluIleMerGlnIleAlaGlnIlyrAlaAspArglysProkAspysle 134
|||||
284397 TTTAAAGTTGTGATATGAGCCATTATGACCAAGAGGTATCCAAAGCAATT 284348
134 uSerGlyIyGlnIyGlnIlyrValThrIleuAlaCysAlaLeuAlaVala 151
|||||
284347 ACTCGAGAGCAAAAGACAAAGAGGTTCCATAGCAAGAGCCTTAGTATAG 284298
151 snProSerValIleuLeuMetAspGluProLeuSerAsnLeuGluAlaIys 167
|||||
284297 AGCCAAAGTCTTTTATCTAGATGAACCTTTAGCAACTTAGATGCTATA 284248
168 LeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValGI 184
|||||
284247 TTTAGATTAAAAATGAGACTTGAAATAGAAAACCTCAACAAAGCTTGG 284198
184 yIleThrIlyrValTYrValThrHisAspGlnIyGlnIyAlaMetAlaIles 201
|||||
284197 AATAACACATTTATCTGTAATCATGACCAAGAAAGATGTTTTCATTTT 284148
201 eTrAspGlnIleAlaValMetIysAspGlyValIleGlnIlnIleGlyArG 217
|||||
284147 CAGATTAAGTAGACGTTTAAATTAAGGGGTATATAGAACATTTGATPACT 284098
218 ProLySGluIleuTYrHisIlyslsProAlaAsnGluPheValAlaThrPheI 234
|||||
284097 CCAGAAAATATTTATTCAAATCCAGACAGAGCTTTGTCAGATTTGT 284048
234 eGlyArgThrAsnIleIleProAlaAsnLeuGluIlyrArgSerAspGlyA 251
|||||
284047 AGCTTTGAAAACCTTATA.....GATTTTAATTAAGGATATCATGATGTA 284004
251 lATyrIleValIlePheSerAspGlyTYrAlaLeuArgMetProAlaLeuAsp 267
|||||
284003 TATATC...AAAAATGAGAGTGGATTAAGTTAAAGTTAAACAAGCTTAAG 283957
268 GlnValGluGluGlnAlaIleHisValSerIleArgProGluIlyrPheI 284
|||||
283956 GATTAACAAATGAA.....GTAAAGCACTATTAAGCCAGATGAT...AT 283916
284 eLYsAspGluSerGlyAspIleGluIyThrIleArgAspSerVal.... 299
|||||
283915 AAGGATTGTAATAAAGAGATTTAGAAAAACACAGTAATGAGTTATAGAG 283866
300 .....TYrLeuGlyLeuAsnThrAspTYrPheIleGluThrGlyPhe 313
|||||
283865 TTAGAACTTTCTTAGAAAGGACATTAACCAATATGATTTAAACAAACCTT 283816
314 AlaSerIlyslleGlnValSerGluIlyrThrPheGluGluIlyrPheGlu 330
|||||
283815 GGGAACTCTAATTTCTAATAGTGAATAATGACAGCATATATGAA..... 283774
330 nLYsGlyAsnArgIleArgLeuArgIleAsnThrGlnIys..... 343
|||||
283773 .AAGGACAAAATATTAACCTCAATTACAGTGATTAATAAATGTAATTT 283725
344 .....LeuAsnIlePheSerAlaAspGlySerGlnAsnLeuIlysls 357
|||||
283724 TATTAACCTTAATATTAATAAATATGTTTATAAAGAGAACTTATAGAAA 283675
seq_name: qb_ba:AE001791

```

```

seq documentation block: 18251 bp DNA linear BCT 02-JUN-1999
LOCUS AE001791
DEFINITION Thermotoga maritima section 103 of 136 of the complete genome.
ACCESSION AE001791 AE000512

```

```

VERSION AE001791.1 GI:4981929
KEYWORDS Thermotoga maritima.
SOURCE Thermotoga maritima.
ORGANISM Bacteria; Thermotogales; Thermotoga.
REFERENCE 1 (bases 1 to 18251)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,M.C., Ketchum,K.A.,
McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
Eisen,J.A., Fraser,C.M. et al.
Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
Nature 399 (6734), 323-329 (1999)
JOURNAL 99287316
MEDLINE 2 (bases 1 to 18251)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,M.C., Ketchum,K.A.,
McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
White,O., Salzberg,S.L., Smith,R.O., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source location/Qualifiers
1..18251
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
92..832
/gene="TM1368"
92..832
/note="TM1368"
/note="similar to GB:Pyro.h percent identity: 70.83;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ABC transporter, ATP-binding protein"
/protein_id="AAD36439.1"
/db_xref="GI:4981930"
/transl_table="MLR1VNLHAKLRDEDEKELKGNVLEIEKGEVHVMGPNKSGKST
LANYMGNPRIYVEGDIVEGNSIKDLPFERAKLGIMTFQNPVEGVKLSQFLI
TAHRIHGEKRYVELEKRELEETKEKLGIDKDFERTLNGFSGEKRKRSIIOSFL
PKRLIILDEIDSGVDVDAIRLARIARLNDEGVTLLITTHYKRLDLHRLKIDKVIHY
VDGRIVSGGPELADIEKEKYSLEGVR"
829..2223
/gene="TM1369"
829..2223
/note="TM1369"
/note="similar to GB:ALD09126 percent identity: 74.45;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAD36440.1"
/db_xref="GI:4981931"
/transl_table="MMERLIITDSDRFNFVRKYATKAPQCLDEKLIMEISRAKDEP
WMLNHRLESLKVENFEMHNPFRGVDISGLDGLKIVSYTKPAKSTSWEEVEEKEAF
DKLGIPEAERKYLAGVAGDSEIYVYONMKELEKMGVIEFLDMSAVREYEDLVKYE
MKLIVPIDHKAFLAHGAIWAGSGTFLYYPAGKIKIMPLOAVFLMSPGQPEHTIIVA
EEGSEVTFIEGCSAPRYNIILHAGMVEIYKKAAYKRYLTIOWMSKNTJNLNKRSTI
VDEGSMTWVSGSISGOKTMYLTPWITIKGCAKARMSYITAGCGHMDPGSKYVHLA
PYTSIYSAKSISLGGGNATFRGLKLTKEVYKSAVECAALMDNRSKSDYPIIE
VETRADVGHGARIGEDDIFILMSRGISEQDAKAMIVKGFVEPVVKELPETAYE
LNKLELEIEKSGIS"
2234..3340
/gene="TM1370"
2234..3340
/gene="TM1370"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1

```

/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAB36441.1"  
/db\_xref="GI:4981933"  
/translation="MEKTKYIIEHDEGKAIVWETPOIFSENDDYEVLKEALENGGPIL  
KEMKKKYOEKEMGEPRMRCNLDGMSLSEFRIEDDVOISLEKLDFEGAHRRKI  
VLLDFFTEPFGVGCDCKHMRFDRGGISENVLYVEKAEIKLSSTGRFPNTWRIRLI  
VKKASAKFRFLDPFGNSFDNFVYLVEGARLVLRDFKAFYRKTHGIIIVLKSSSK  
ADMSFTFYQNSGISITDLALIPRLREGSDARGLKGNVVNDGAQ IYVRFI LMYTRKRSK  
IAVETSTYGVSGTDAAMDIAIPMLVDENDDVAISHASSSSSLDBELFYLTVKRKSK  
EKKR.IYRGVFNEMLDELDESVEGVENVNRIY"  
3318..4559  
/gene="TMJ371"  
3318..4559  
/gene="TMJ371"  
/note="similar to PID:1001734 PID:1001708 percent  
identity: 61.37; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="aminotransferase, class v"  
/protein\_id="AAB36460.1"  
/db\_xref="GI:4981951"  
/translation="MLSTGYEDFPPILKKINKGRIVYLDASAATTLKPKOVVKLEE  
YNSYSAHVRAVHTLAIVARTTEEPREAREPAGTSPEIIIFTGTMAINLAYSIL  
LRSLIEDVLSLEEHANVPWLRLSKFHGRTIDTFRSGRGFLEMDDLKHNR  
DKNRVVALTGLSNVTGQRIPVELKNVFNPSIILLDGQOLI.PHEPVKPKKEIGVPLA  
FSLMKMGPRGVGLYCKRREVLEOMEPFLYGGMDIRYSLEDYFENEIPLYFEAGTP  
IADIVASREALRYLKETGSPDSVHEKI EOLTOMALKPMKIGGVETIYGLDERHGSIYA  
FNWGHIHPDHVAHLHOEFQIANRSHHCQAOLMSLTLRKQSIDPENSTCASPFYIN  
TEEDVKLTUESGKYKVKMFESR"  
4544..4972  
/gene="TMJ372"  
4544..4972  
/gene="TMJ372"  
/note="similar to GB:U00013 PTD:466875 PID:2398702 percent  
identity: 55.07; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="nitrogen fixation protein NifU-related protein"  
/protein\_id="AAB56442.1"  
/db\_xref="GI:4981933"  
/translation="MPFKMYSEAIIIDVANSKKERKGLDADVIEEGNISCGEITTL  
LYLKVEDVVDNAKREGSCVYISQASLSMLERIGERYEIRSLTEEAEMKSNGENFEP  
EGKLKTNYLMSDIKDINYAPARKCFCLAMTKTLKALKISR"  
complement(4950..5936)  
/gene="TMJ373"  
complement(4950..5936)  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAB56443.1"  
/db\_xref="GI:4981934"  
/translation="MRKLSIVFLIVALSVMGSHEGLTYIVKDIVPDENVITTPY  
NYGNRVNYLSPKILINGDCOITIASMBGALII.PNPSPEDGKVPMOLLITYSSBPM  
CMGDGLDLPLOPILENGSOVRRIMKRTIGGLEFFPADKSFYIFHDMAEKARRKRDGRI  
GYRLDALALHYIELDSOPHNNTGEGEVLETFOALPDSIKMKIMNANHAYDYDLAVLYL  
SGDEVIKTILEADPAVLKSEHMLVDEVKRLASYPRKYHRKIKVFOEKLKERISDPI  
ESPFIYVEEFREDGDLVLKRYTLBELICKVSSYLGXYLIDLAKAVKSF"  
5984..6589  
/gene="TMJ374"  
5984..6589  
/gene="TMJ374"  
/note="similar to GB:AEB00657 percent identity: 62.38;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoglycerate mutase"  
/protein\_id="AAB36444.1"

	gene	/db_xref="GI:4981935"	/translacion-"MKKYLIRHGETINNEGNGVTDVLNENGRQARKLANSIKR VDATISPLKRSETAEIARREKETIVEDRECEI SLANGLTVEALIRETVERK KWSDPNMGEGLESMRNVRVAKIMKTIVSOEKLGSNNYIVSHLSLRATFCWI LGGLPYLRHNFKEDNASLSVYEIESKPRFLVILNDTCHLKES" 6627..6700
	CDS	/note="tRNA-Cys-1"	/product="tRNA-Cys"
	gene	6627..6700	/gene="tRNA-Cys-1"
	gene	6756..7766	/note="similar to GB:M64519 SP:P23861 PID:147329 GB:U00096 PID:1651550 percent identity: 60.95; identified by sequence similarity: putative"
	CDS	6756..7766	/codon_start=1
	gene	/transl_table=1	/product="spermidine/putrescine ABC transporter, periplasmic spermidine/putrescine-binding protein"
	CDS	/protein_id="A0D36445.1"	/db_xref="GI:4981936"
	gene	/translacion="MMKKYLVLLVLAIFPAKATLYINMADYPEDVIKRAFEKY DCRVYDTYSNEEMAKFRAGGCKGDILFPSSDYISIKKBGMLOKDLISKIPNK YLKDILAKTYDPNHETSVPMKGSTVIVNKKYVKEYSWIFPREDLRGMTLL DDMEHVGAAALKTYGYSVNTNRPELEBAQOVLMRMKNIAKPDASSADGIYSGEYW VHGADVADYORIPGECEKEYFEEFIPEKGCTLMNDNVIPRGANVDAVKFINFILE PENNAKIDYVIGLSPVPEAKRYMQTEPIYTIEDLNKCEFIEDYGALDELNYINKTWELI IM"	
	CDS	7781..8887	/note="similar to GB:M64519 SP:P23858 PID:147326 GB:U00096 PID:1651555 percent identity: 67.38; identified by sequence similarity: putative"
	gene	/transl_table=1	/product="spermidine/putrescine ABC transporter, ATP-binding protein"
	CDS	/protein_id="A0D36446.1"	/db_xref="GI:4981937"
	gene	/translacion="MIGEVSIRKNSKFDDFOYLNKNSLDIKKEPFSLGPSGCK TTTLRIAGFEVSGCVLLDGKSILMLPNKRPNVNIIPONVALFPALTVEENIAFPL KIKKLSSEINORYNELLSLIRMEHQPTTPHSDQAFAISVSRVRLAMNEGTVYLD EPISALDAKIQRELIVELDNLDHREVGTTFYVTHDQAFASVDRVRLAMEGETVOG TPEEVESPVNVERPATFTGETNLKKAIVEVEDITYVESPGIGOPCYTRDKERKDD RLITLTPKERIRISRKQFSDEETFNFGVADERIYGHOTKYFRULDESYIMKVYKQ HARYIIDPELIKWEDEVFITWNPDSFIVELEE"	
	alignment_scores:	Quality: 701.00 Length: 327 Gaps: 4	Ratio: 2.804 Percent Identity: 42.202
	Percent Similarity:	76.453	Percent Identity: 42.202
	alignment_block:	US-09-769-787-162 x AE001791 ..	
	Align seg 1/1 to: AE001791 from: 1 to: 18251		
	3 GUUIELYSIIIEIASNALAYSLYSILEYRHISAPVALPROVALI 19		
	:         :     :     :		
	7793 GAAGTTTCATAAAGAAGACCTCAGCAATTCTTCGATGATTCACAGTTT 7842		
	19 eGIuAsnLeuAsnIleThrlIeProLysGlySerLeuPheThrLeuLeug 36		
	:::     :           :     :     :		
	7843 GAAGAAGCGTTCCCTGGATATAAGAAAGAGAGAGTTTCTCCATACATCG 7892		
	36 lyAlSeGlgYsgLyStHrTrHleuLeuArgMetIleLaGlyPhe 52		
	:     :     :      :     :     :		
	7893 GCCCATTCGGGTTTGGAAGAACACGCTTCCTTCGCTGTACTACGGGGCTTC 7942		

```

53 AsnSerIleGluGlyGluPheTyPheAspThrLysIleAsnAs 69
   :::::::::::::::::::::
7943 GAGGGGTGGAAAGTGGACATGTTCTCCGATGGAAAGTCACTGAA 7992
   :::::::::::::::::::::
69 nMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnTyraIaI 86
   :::::::::::::::::::::
7993 TTGGCTCGGAACAAAGCCCGGTCAACATCATATTTCAGAACTACGCTC 8042
   :::::::::::::::::::::
86 LePheProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGln 102
   :::::::::::::::::::::
8043 TGTTCCTCATCTCACAGTTTGGAAACATCGCTTCCCTGAAACTC 8092
   :::::::::::::::::::::
103 LysLysValProLysGluGluLeuIleGlnIleThrAsnLysTyreLueG 119
   :::::::::::::::::::::
8093 AAAAAAATCTCCGAGACCAAAATTAATCAGAGATCAAGCAACTGCTTTC 8142
   :::::::::::::::::::::
119 uLeuMetGlnIleAlaGlnTyraIAspArgLysProAspLysLeuSerG 136
   :::::::::::::::::::::
8143 GCTCATAAAGATGAAAGAACACGCTCAGAAATGCTCCCAACTGCTCG 8192
   :::::::::::::::::::::
136 LysGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnPro 152
   :::::::::::::::::::::
8193 GTGGTCAGAGAGAGAGTGGCGATGAGCGGCTCCGCAACGAGCGG 8242
   :::::::::::::::::::::
153 SerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuAr 169
   :::::::::::::::::::::
8243 AGGGTCTCTCTCTCGATGACCGCTGAGCGCTCTCGACCAAACTGAG 8292
   :::::::::::::::::::::
169 gluAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleT 186
   :::::::::::::::::::::
8293 ACAGCAACTCTTGTGAGCTGCACAACTTCATGACAGAGTGGATCA 8342
   :::::::::::::::::::::
186 hThrValLysValThrHisAspGlnGluAlaMetAlaIleSerAsp 202
   :::::::::::::::::::::
8343 CCTTCATTACGTAAACACAGATCAGCTGAACCAATCAGTGTTCAGAC 8392
   :::::::::::::::::::::
203 GlnIleAlaValMetLysAspGlyValIleGlnIleGlyArgProLy 219
   :::::::::::::::::::::
8393 AGAGTCCCTCATGACGACGAGAAATCGTTCAGGTGGAACTCCCTA 8442
   :::::::::::::::::::::
219 sGluLeuTyRHisLysProAlaAsnGluPheValAlaThrPheIleGlyA 236
   :::::::::::::::::::::
8443 CAGAGTGTATGAGAGCCCGGTCAACGCTTTCGACGACCTTCATCGAG 8492
   :::::::::::::::::::::
236 rGThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTy 252
   :::::::::::::::::::::
8493 AAACGAACTCATGAAAGCCGAGGTGTTGAAGTGAAGTGAATATTTAC 8542
   :::::::::::::::::::::
253 IleValPheSerAspGlyTyraIAlaLeuArgMetProAlaLeuAspGln 269
   :::::::::::::::::::::
8543 GTCGTGAAGAAATCCGGGAATCCGCCAGTTCAGGTTCACAGAGACAAGA 8592
   :::::::::::::::::::::
269 lGluGluGlnAla.....IleHisValSerIleArgProGlu..... 281
   :::::::::::::::::::::
8593 ACCAAGAAGGAGTATGATGCTCATCACTTCAGACCCGAAAGATCA 8642
   :::::::::::::::::::::
282 .....GluPheIleLysAspLysSerLysPle.....Glu 292
   :::::::::::::::::::::
8643 GAATCTCAAGAAACAGTTCTGTCGAGAAACCTTCATGATTTTCAC 8692
   :::::::::::::::::::::
293 GlyThrIleLeaArgSerValTyreLueGlyLeuAsnThrAspTyPheI 309
   :::::::::::::::::::::
8693 GAGGTGTGGATGAGCAATATATCATGGGCGACCAAGCAAGTACTTCGT 8742
   :::::::::::::::::::::
309 e.....GluThrGlyPheAlaSerLysIle 317
   :::::::::::::::::::::
8743 TCGACTCGAGCAAGCGCTACATATGAAGTT 8773
   :::::::::::::::::::::
seq_name: gb_ba:AF307052
seq_documentation_block:
LOCUS AF307052 18001 bp DNA linear BCT 12-FeB-2001

```

```

DEFINITION
Pyrococcus furiosus transposase, sugar kinase, trehalose/maltose
binding protein (malE), trehalose/maltose transport inner membrane
protein (malP), trehalose/maltose transport inner membrane protein
(malG), putative trehalose synthase, trehalose/maltose transport
ATP-hydrolyzing protein, putative sulfate transport system permease
protein, putative sulfate transport integral membrane protein,
putative sugar-binding transport ATP-binding protein, putative
solute binding lipoprotein, and transposase genes, complete cds;
and unknown genes.
ACCESSION
AF307052
VERSION
AF307052.1 GI:12018039
SOURCE
Pyrococcus furiosus.
ORGANISM
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1 (bases 1 to 18001)
Diruggiero,J., Dunn,D., Maeder,D.L., Holley-Shanks,R., Chataud,J.,
Horlacher,R., Robb,F.T., Boos,W. and Weiss,R.B.
Evidence of recent lateral gene transfer among hyperthermophilic
archaea
Mol. Microbiol. 38 (4), 684-693 (2000)
JOURNAL
2056786
MEDLINE
11115105
PUBMED
2 (bases 1 to 18001)
REFERENCE
Dunn,D.M.
Direct Submission
Submitted (20-SEP-2000) Human Genetics, University of Utah, Rm 308,
BPRB, 20 S. 2030 E., Salt Lake City, UT 84112, USA
FEATURES
source
1..18001
/organism="Pyrococcus furiosus"
/db_xref="taxon:2261"
60..840
/organism="Pyrococcus furiosus"
/insertion_seq="IS-like element"
84..833
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="AAG45369.1"
/db_xref="GI:12018040"
/translation="MSGISWYQYRRKRGNNKSPETIIYVVSALKPPRRNKIPPEKTI
RGVELYLRGLSYROTARILKLSHTYWEAVOKLAAYKKRLLVYKQORNIANDRY
IKIKGKRYLEMAIDVESKEVLAIVITVRKMWARQDILVLSGCRQPFVLDKRS
WYKSAFSLRGLYLVTFEGPNSVSRWFRLKERTKRPWNFRGKDWRRVHRFVLEFA
FWYNFVRHSFSDPPGSDYEMLEVMPLS"
903..1220
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45370.1"
/db_xref="GI:12018041"
/translation="MKVAVVYLPLFOSFLVYSARGSERVMLIFGLNPLFAMGI
VKDRRAKTIATKGYDFFALLMNGALIFQSLNAIDLAIIHDLGLGKKEPPT
EES"
complement(1204..2142)
/codon_start=1
/transl_table=11
/product="sugar kinase"
/protein_id="AAG45371.1"
/db_xref="GI:12018042"
/translation="MIYATGELIDFLAKEGCKIKDVREEFKHPGAPANVVGRLRL
GAKSALISKVGDDEPGEFLIEELKREYETVYIIKDNKRGIVFVOLIAGKPEFTLY
DGVAVFNLREKEIOWDEMDALDELHFGSVLARPSSSTVEVLRVAVGKVPISYDVN
IRLDLWGREREKEMIDIEALADIVYIGDELEYLNKNGIALEDFEFALVALTRGA
EGSFTIHKDIDVDVPSFVPEVDVDTGADAPMAALIASLFYNGKLDLIEFSKEELKEL
GSFANLVAALSTTKRGANSVPSLEEVLRKRFSLP"
2376..3728
/gene="male"
2376..3728
CDS
gene
CDS

```

```

/gene="male"
/codon_start=1
/transl_table=11
/product="trehalose/maltose binding protein"
/protein_id="AAG45372.1"
/db_xref="GI:12018043"
/translation="MKKKVVLGLFLVGVGLAVVASCIGCGQISTVSTPETSLO
GKIVEAVGAPNEIEYWKVIAEFKRYGVYELKROATDTEQRDLVNLKRSKSS
DDVPLMDVAMLGQFVASCGLPELDYVORNDVLSVFPQSYINLADKGGKLYALP
YIDAGLLYRKDLLEKYGSKPPEIWOELKMAOKIIOGSERETNPNGEVMQOKYE
GLVDEVEYVYVNGSGISGEFKDGKVPITLKPENVEALQPMVDLHKYKISPNYTE
MTEPRALMEQOGNAFERNWPAMGLAHNADSPYKGVAPLPHFPHKSKATLGG
WHIGISKSDNKALAMEVKEVSEISVOKGFAMNLGMPGRVDYDDPAVYKSPULK
ELRAVENAVPRIVYPYQPSLSEIIOKYVNSALAGKISPOEALDKAKOELELVKQYS
"
3770..4663
/gene="malF"
3770..4663
/gene="malF"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane
protein"
/protein_id="AAG45373.1"
/db_xref="GI:12018044"
/translation="MDNNLTSLKLYREAKLGYLMLPLTTVVLFFILPMGTFWISL
HBDVPIPEKPEVGLRNLVRLVLSAREFVSTFVSFVSYSILGLSPALINE
RLKRGVLAIVLIIPWAVPTIISARFEMELWYSLGFNMWISLIGSVNMLGPIS
AFPAIVADWKTTPPEMTLLIAGLOAIPODLYEALIDGASMEERKSTPLPLKPV
LVALLRTIDALRPFDITIVLTGGSGGATTSISLARNTYNLGDYIGSAISLTF
VVLSTIYVVLVGRREGSK"
4660..5496
/gene="malG"
4660..5496
/gene="malG"
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport inner membrane
protein"
/protein_id="AAG45374.1"
/db_xref="GI:12018045"
/translation="MREYVKRLILITGAILMAITLFPFIMIVSPADPTFGSP
LVEYKSTLENYRVLSDPYLHPFAYLKNSTIIASLTTLTTSISSIAAASRIEKG
RLILPIFLVGLSPQISLVGLFKFIKLGWNTYQALYFVAWATPLSLMILTSLY
FSQLPDLDEAAMIDGASRIKLTITLPLSAPALSTALVFIANFENFMALLFTT
DHRARTVPGIALFGVGEIPIWGSYMAVSISTPIVLIMALLFKYIVSGLTAGALK
GE"
5501..6739
/codon_start=1
/transl_table=11
/product="putative trehalose synthase"
/protein_id="AAG45375.1"
/db_xref="GI:12018046"
/translation="MEYVYKFGSGEKRLDYEIIGDEALEAKAKAENLQKSPYHV
NSTSGEGVAILHNLPMLRDVGDARNFVLEGNEEFNVKSPFNALQKNEELKT
EEMKRLYLIEINKNAMEDIDLQFDVYLHDPPAPLIEFEKROPIMVCHILDSPN
LEFMKLRQFVEKRYRIFHMEYVOEDLNOKRVVIMPSIPLSKNNELSESLST
TLEREDVDEPRLITOVARFDEPMKGVEDIVDYRKKEKIPPEOLLGVMAHDEEG
WYFEKTKTIGEDDIDKVLNTLGVHAREVNAFOASDVYIIQMSIRGEFGIIVIEAM
MKKEPYGKAVGSGIKIQIVDGKTGPLYKVNDAIENTTLYLLEHKDYAODGMKNAKRI
KENFTIHKLERLIDLNSF"
6772..7788
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45376.1"
/db_xref="GI:12018047"
/translation="MEIPEISHALSELIGFTRKYEITLTYTLVVGSPSTAKEISTKSGI
PYNRVYDITSLKRGFVEIGTPKVAAYSPRIAFPEKKELEIMKLEILNVA
KKEORPAIMRSRDEATFEMPRESLYSAKNVIYVTPPEPFEITREDILKLEGGVT
VSLYIDKIPDLSEFKGKNFVROPIKLNHLGMDGKVVYVITQNTPTPSISPPSKS
TYPEITFSOYSLIETFEKSTLEKETTIGPKDIRFPAEMHADVFNKHLKNNIYAEI

```

```

TCKNLESGRLFTLGRVVGTYTSLREAVNNHILETENGVYKVGMPAVIEDYESTEIK
FMG"
7794..8912
/codon_start=1
/transl_table=11
/product="trehalose/maltose transport ATP-hydrolyzing
protein"
/protein_id="AAG45377.1"
/db_xref="GI:12018048"
/translation="MAGVRLVDMKVEGTVAVRELSLEVKDGEFMILLGPSGCKTT
TIRMATAGLEEPSRGQIYIGDRLVADPEKGIIVPAPDRDIAMVFOSYALPHMTVDNI
APFLKRYPROEITQORREAVLELIGTELLNRKRELSSGGRQVALAGRAVRRPOV
FLMDEPLSLDKLRLVRMAKLELKQROLGVYTIIVHDQVEAMTGDITAMNGVL
QOOGSPDEYVDKPAFTVAGFISGSPMNLDAIVDEGDVDEGEERLKLIPQEFVIG
ELGVGREVIFGIRPEDLYDAMFARVARGENLVAVEIYVNLGSEIRIVHLRGVGT
FYAFRSRVRVEGVVDVFDMKKIHIFDKYTKGAIF"
8959..10461
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45378.1"
/db_xref="GI:12018049"
/translation="MRMCVLAVTTFDPRTALSVREKALKKKHNLKELKSSN
LEVFVNEALGRYEALERGENGVVDKDEEVLAKSLISRDISGIIILGIMHTESNLV
TLARETNKPLLVDADDDPAMAGSTCITSVGSALNESSVNYHALHHVRLKGVREKVA
MARAAVSKLSKSLLMGAYTTGMEHMDLRLKRFIDGFLIDQYTIKRRADS
MSDKERRRVEFEFDMENKAKVNDKMLPEVLRQIAIYLAOKIMREYBOETV
AVSICQPELSIFCYVTACLIPALRPNADAKGEKPYVAPDCEGVKGTISSALLFYL
SGKPLFGDIKYVDELVYIACGASSLYAKLSNPENNELKETTIOGOCOKSGSAL
TYRPTPKLTVARLIRKGEYLLFLAEGVITEIISKLEKLMGWQWMTALRNPLDK
EKFNLIMGANHLSAVPDTEIEFEFVARWMIKAVNLKKEVENFNTLSL"
10486..12399
/feature="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG45379.1"
/db_xref="GI:12018050"
/translation="MSVILSYNGAFVARTRONGDMKDGXYFIPTDRFLKGVKLKDK
NSVLIGSODGPRKAYSHSIEDVAVLTKRKRIKXMAVRELOFYNYSKRHVLRVE
YSFKPIEDIEFVRKFGGOKIRRIKSSLDGDNESYTGKOKIRKNLIKIKTMREKG
LAFATITTEPLEKEFLYLFETIKISKEGELITLTKPLLPVAVSTNLMDRVERA
IEDLUALATYVYGVAPGAPLRYVACPEGRDSIITSWLLPYYPQYAGCTLKFEAKIQ
GKKFPLNDEEBEGKIDHREFGELSLSHKMPAPRYGGLAONGMDSKEGIPTEGPTPK
RELIRLAKNPLAAVEMLIRLDEGGVIRYGGGLAONGMDSKEGIPTEGPTPK
PVALYVGVYAKKALMDAASIGLTLSDPRMLKKEAEELKRRNKDFCWDGFALADG
DNVPSKYVSNNGHILFTGIAHOKIAERLFEEDMPSGMCIRTLSKREKAYNPSYH
NGSTWPHDAVTAIGLSTIGEKERAMLAELFTAKFLPNSQLPELRSGLSEYPL
CPRANAPOQMAASAVFALTALLGLKAERELVSPLRNFKVSALVBRFRKAYLIES
QKEEVYRFEERDI"
12350..13339
/feature="conserved hypothetical protein"

```

alignment\_scores:      quality:      length:      334  
                         ratio:      gaps:      7  
Percent Similarity:    67.665    Percent Identity:    44.611

alignment\_block:  
US-09-769-787-162 x AF307052 ..

Align seg 1/1 to: AF307052 from: 1 to: 18001

```

15 AspValProValIleGluAsnLeuAsnIleThrIleProlysglyserle 31
||||:|||||
15043 GATATTCCTCTCCAGAG.....GCCGAACCGGAGAGGTT 15077
||||:|||||
31 upheThrLeuLeuGluAlaSerGlyCysGlyLysThrThrLeuLeuArgm 48
||||:|||||
15078 TTTACTCTGCTGCGCCGAGTGCTGTGGAAGACAACTTGA 15127
||||:|||||
48 etIleAlaGlyPheAsnSerIleGluGlyIupheTyPheAspAsp 64
||||:|||||

```

```

15128 TAAATGCTGCTTTGAGAGCCAGCAAGGAGATATACCTTATGAT 15177
15129 ThrLysIleAsnAsnMetGluProSerLysArgAsnIleGlyMetVal 81
15178 ACCGTTATGAACGATGCTCTCCATATGAAGACATAGGATGCTCTT 15227
81 eGlnAsnTyrAlaIlePheProHisLeuThrValArgAspAsnValAla 98
15228 TCAGATATATGCTCTCTTCCACATATGACGTTATATGCAACATCTCT 15277
98 heGlyLeuMetGlnLysLysValProLysGluGluLeuIleGlnThr 114
15278 TCGGTTTAAAGCTCAGAAAGCTCTCTAAGAGAGATTAAAGACGAGTC 15327
115 AsnLysTyrLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLys 131
15328 TCATGGGCTCTTGAGCTTGTTGGTTTGAAGGATTTGAAGACCTTACC 15377
131 AspLysLeuSerGlyGlyGlnGlnArgValThrIleAlaCysAla 148
15378 CGAGCAGCTGACGCGGTGTCACACAGCAGAGATTGCACTGGCAGGCTT 15427
148 euaIaValAsnProSerValLeuLeuMetAspGluProLeuSerAsnLeu 164
15428 TCGTATAGAGCCCGACGCTTGTGCTTTAGATGAACCTCTTCCAACTTC 15477
165 GluAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGlnIleGln 181
15478 GATGGAAAAATATAGAAAGAGCTTAGGGGAGAAATTAAGATCCAAAA 15527
181 sGluValGlyIleThrValTyrValThrHisAspGlnGluAla 198
15528 AAGAGCTTGAATACATCAATCTATGATACGCATGTCAGGAGAACTTA 15577
198 eAlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGln 214
15578 TGCACATAGTAGTAGATAGACCGCTTATGAGCGTTGGAAGATAGACAG 15627
215 IleGlyArgProLysGluLeuTyrHisLysProAlaAsnGluPheAla 231
15628 GTTGGAAACCATGTGATCTCTACCAACCAAAAAAATGAGTGTGTAC 15677
231 aThrPheIleGlyArgThrAsnIleIleProAlaAsnLeuGluLysArg 248
15678 CAGGCTTTTAGCTCTTCTAATCTTTAGAGGTTAAACGACAGAGGAA 15727
248 eAspGlyAlaTyrIle.....ValPheSerAspGly 258
15728 AA.....GCATATGTTGGAACTTTCCTTGAAGTTAAACGAGAGG 15771
259 TyrAlaLeuArgMetProAlaLeuAspGlnValGluGlnAlaIle 275
15772 .....AGACTAA 15779
275 sValSerIleArgProGluGluPheIleLysAspGluSerGlyAspIle 292
15780 GATATCTTTAGCCCGAGAGC...GTTATATCGAAGAGGAGAGC.... 15822
292 LucGlyThrIleArgAspSerValTyrLeuGlyLeuAsnThrAspTyrPhe 308
15823 .....ATGGAGAAATAGTGCACATCGAA 15846
309 IleGluThrGlyPheAlaSerLysIleGlnValSerGluGluSerThrPh 325
15847 CTGCTCCCGGCT..... 15858
325 eGluGluAspLeuGlnLysGlyAsnArgIleArgLeuArgIleAsnThrG 342
15859 .....AGAAATACGCGCTCAAGATTGATGTGG 15883
342 In 342
::

```

```

15884 AG 15885
seq_name: gb_ba:AF307053
seq_documentation_block:
LOCUS AF307053 17854 bp DNA linear BCT 12-FEB-2001
DEFINITION Thermococcus litoralis sugar kinase, trehalose/maltose binding protein (male), trehalose/maltose transport inner membrane protein (malG), putative trehalose synthase, trehalose/maltose transport ATP-hydrolyzing protein, putative sulfate transport system permease putative, putative sugar-binding transport ATP-binding protein, and unknown genes.
ACCESSION AF307053
VERSION AF307053.1 GI:12018057
KEYWORDS
SOURCE
ORGANISM Thermococcus litoralis.
Thermococcus litoralis
Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
REFERENCE 1 (bases 1 to 17854)
AUTHORS Diruggiero, D., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F.T., Boos, W. and Weiss, R.B.
TITLE Evidence of recent lateral gene transfer among hyperthermophilic archaea
JOURN. Mol. Microbiol. 38 (4), 684-693 (2000)
MEDLINE 20566786
PUBMED 11115105
REFERENCE 2 (bases 1 to 17854)
AUTHORS Dunn, D.M.
TITLE Direct Submission
JOURN. Submitted (20-SEP-2000) Human Genetics, University of Utah, Rm 308, BRB, 20 S. 2030 E., Salt Lake City, UT 84112, USA
FEATURES
source
1..17854
/organism="Thermococcus litoralis"
/db_xref="taxon:2265"
844..1161
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=1
/product="sugar kinase"
/protein_id="AAG45387.1"
/db_xref="GI:12018059"
/transl_table="MiyaiGEIILDFIAKEGKLKDYREFEKGAPANVVGRLGKSAISKVGDOPGEFIELEKKEVERKRIIKDNKHTGIVFOLGAKPEFITYDGVAIVNLKREKELQMDPMDAELHRSYVFAFRPSSTVEVRAVKGVPISYDNIIRDLWRGKREKELDIEDALKLADIVKIDGELYLNKGIATEDNPAVALITRGAEGSTIIHKDIRVDVSPKVEVDITGADAFMALLSLFYMGKIDILLESKEELKELGSFANLVAAALSTTRKGAWSVPSLEEVILKRRFSLP"
2317..3669
/gene="male"
2317..3669
/gene="male"
/codon_start=1
/transl_table=1
/product="trehalose/maltose binding protein"
/protein_id="AAG45388.1"
/db_xref="GI:12018060"
/transl_table="MNYKRYVLGLFLVGYGIAVYASGCTGGQOTSTVSTPIETSLQKTIYAVGAPANEIEYWKVYIAEFERKYPGVTEYLKQATDTEGRRLDLNALNGKSSDPDVFMDVAMVLEQFIASGWLPEDDVVDKNDLSYFQSVINLADQCGKLYALDVIYDAGLLYRKDLLEKYGYSKPEPMQELVEMAKIQSGERETPNPNCWGVMOGKQYE

```

gene  
3711..4603  
/gene="malp"  
CDS  
3711..4598  
/gene="malp"  
/codon\_start=1  
/transl\_table=11  
/product="trehalose/maltose transport inner membrane protein"  
/protein\_id="AAG45389.1"  
/db\_xref="GI:12018061"  
/translation="MDNNLTSLKRYAKGLGMLPLPLTVLVPIILPVGMTWISLHSDVEIPEKPEFVGRNLYRLVLSAREFWYSTFVSPFSVLETTILGSPALILNEFLKGGVLRATVILIPMAVPTIISARTWELMKNYSGLFNMTISITIGSPVNMGLGISAFRAIVADVKKTTPPTMLLLAGLQALIPQDILEKALIDGASMEERKSTIDPLKPLVLIVALILFTIDALRFVDIIVYLTGGPGGATTSISLLAFNPNIGDVGIGSAISILTFVVLSTFIVYLRKGRFRD"  
4600..5436  
/gene="malg"  
4600..5436  
/gene="malg"  
CDS  
4600..5436  
/gene="malg"  
/codon\_start=1  
/transl\_table=11  
/product="trehalose/maltose transport inner membrane protein"  
/protein\_id="AAG45390.1"  
/db\_xref="GI:12018062"  
/translation="WREEVLRKRLIILIGAILMAIICLPETIMTVISFAEDPPLGSPLVKYSTLENTVNRVLSPTLHPAYLKNISIIASLVLTIVSISLAAYAVSRLEFGRLILPIEVLGSMFPOISLVGLFEKIEKLGWNYQALFPYVWMTPLSLMILSYFSQAPKDLDEAMIDGASRIKTLTIIPLSAPALFSPALLVFTAAFEPMFALFTTDHRAATVPGVIALNQGVCGLPWSVMAASVISTIPLYMALLRQKIYVSGLTGALKGE"  
4641..6679  
/codon\_start=1  
/transl\_table=11  
/product="putative trehalose synthase"  
/protein\_id="AAG45391.1"  
/db\_xref="GI:12018063"  
/translation="WREYTRFGGSGKRLIEDYREIIGDPALEIAIKAKANLKDQSVVHNSISFGGVAEILHLVPLMRDVGIDAMFVIEGNEFENYTKSFHNLQGNKELRLTEMKLYLEIKKNKNAEDIDLQDFVYLIDHPQAPALIEFYERQPMIRHLDIDSDNLEFMKFLROFVEKEDRYIFHMEEYVOEDLNQEKVYIMPSIDIPSEKMEISESILKTLERDVDPERPIITOVARPDPMKGVPRVIDYVYRKVEKIEPEVOLLIGVMAHDDPEGMIYEKTLRKIGEDDIKVLNLTGVAHAREVNAQRASDYLQMSIRGPGCITYTEAMWKEPVGRAVGGIKLQIVDGRTGFLVADVNDALIKTLVYLLEHKDVAOEMKNAKERIKENFIIYKHLERYLDLNSF"  
6712..7728  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAG45392.1"  
/db\_xref="GI:12018064"  
/translation="MEIPEIISHALSEIGFTKYEITLWTLLVYGPSTAKESITRSGIPYNRVDTISSIKLRGVEITIEGPKPKVAASPRAPFRFKELEDIMKKEILENNYKKEORAPAIHRSRSDPALEIMPRBSIYAKNEVIVYVSEFETIRELLITLLEGVTVSLYIDKIPDLSEERKGNFTVROBTYKLNHILIGMTGEEVATIQATIDSGIPBSFKTYPEITFSQYSLIIEIFKSTLEKEIIGNPFIAPFAVDPMFKNLKRNRYAEITGKNLSGRLETLTGRVVGTYLSLREAVNNHILTEENGVCVVGWFAVIEDYESTEIKFIMG"  
7734..8852  
/codon\_start=1  
/transl\_table=11  
/product="trehalose/maltose transport ATP-hydrolyzing protein"  
/protein\_id="AAG45393.1"  
/db\_xref="GI:12018065"  
/translation="MAGVRLVDVWKVGEVTAVALRSLLEVADGEPMILGSGGCKIT

TLRIAGLEEPSRCQIYIGRLVADPEKGIIVPPKRDIDIANVFOVALYPMYVYDNI  
AFPLKLRKVPQIDQREVAELIGLTELNRPRRELSGQQRORVALGRAIVRKPOV  
FLMEBEPINLOAEIKRIRKIKSSLSGSENYSTGAKIRNKLTKNMTEG  
QOVSPPDEVYKPAINTFVAGFISGSPFMFLAITYEDQFVGFGRLLKLLPQDEVLG  
ELGVGDEVYKIRPREDYDAMPARVRPQGNILRAVEIYVENLSGSERIIVLRGCVT  
FVGAFRSRSRRVREGVEVDVFDMKKIHFDKTTGKALF"  
8899..10401  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAG45394.1"  
/db\_xref="GI:12018066"  
/translation="MPWCVLAVITFTDPKRTALSYEREKALLEKHNLIKELKSSN  
LEVPDVALEKLYALERGENGVDDKEVYKASRLIGSRDISGIIIGLMTESNLV  
TLAKETNRKPLLIADDDPAMAGSTCITISVAGSLMESSVNYHALHYRLKGDVEKVKY  
MARAEAVSKLSKSLSLGMAPYTLGMEHLDDPLPKRFRIGDFLILDOQYIIKRA  
MISPEKRIIVYEEPPDMLNEKAKYNNQMKMLTPVLRQIAYLAADIKMEYQDEY  
AVSTKQPELEIYGVTRCLIPALPPRPAADKGEKPVYPATCEBDYKGTISSALLFYL  
SGKPLFEDIDYVDDELVLNCGASSLYAKLSNPEENLKEETTGOGCGKSGGAL  
TYRTPPTKLVARLIRKGEYLYFLAEGEYITEEISESKLWKGRQMPHTVIRNPLDK  
EKLINMGANHLASVPGDFTIEIFVARMGIRKAVNLRKKEVENFLTSL"  
10426..12339  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAG45395.1"  
/db\_xref="GI:12018067"  
/translation="MSVILSYNGAFVATRONGDMDKGDGFIYEDTRPLKGVRLKLOK  
NSVILGSDQCPKRAKYSHFSIEDVYLIRKEITKDNNAVRELOPYNTSKRYALRE  
YSFVPIEDIEVRRFGQKIRRKIKSSLSGSENYSTGAKIRNKLTKNMTEG  
LAFAPILIEPLEKLETLLEFIPKISKBELEFLEKPLLPNNVSTMLTWLDVFERA  
IEDTALTAIYNYGMVFPAGIIPYACFGSRDITSMFLPYYQVAGEGLKPFKAIQ  
GKNNLMEEPGKIIPHEFRGELSHSKMPFAYYGVGDATPLVYLAAYLRMTGD  
RELIERKPNLTAAVEMTLRLDGGGYIRKGGILNAGCKDSKEGIPREGPTP  
PVALYEVGYAKALMDAASGLTSLDPKMLKRABELKRFKNDPCDGYAALDQ  
DNVPSKVVSSNMGHLFTFGIAEHQEKIAERLEEDMFGSGIRTLSSKERAYNPFSTH  
NGSIMPNDNAVIAIGLASIGERKARMAEAIFFTAFLPLPSQLPELFGSLESEYPLL  
CPRANAPQWASASVFLATLALGKAEKELTVPLLPKNKVSALVFRKAYLIES  
QEKERVPEERDI"  
12290..13279  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAG45396.1"  
/db\_xref="GI:12018068"  
/translation="MNLRRLDLKSLFPIKIKIGSNKIDIKPKKIPOPILSPSR  
EGFDSRNTYNPVAVIKEDRIYMLYRAESKEKTLGRGLAISTDGINFPRPEITIEP  
EYKWSVEVDPRIVKRGITYMYTYTGDTARLCLATGATSNMLTKMHGPIFGFEY  
KKNNIKGTWKSGLALPKLKEKNGFNGLMFGSNNIMAVSKDLVNMVEYKKEVLSR  
REGEDVNLVPRGPLLETSTYGLIAYSAGRYDKLKYKGVVIFPKREYDKVAART  
ETPLMVPEYEMELYGHVNNVYFVGWHEHKKILLYYGADRHIGLAWYTDL"  
13303..14934  
/note="abc transporter"  
/codon\_start=1  
/transl\_table=11  
/product="putative sulfate transport system permease

alignment\_scores:  
Quality: 683.50 Length: 334  
Ratio: 3.024 Gaps: 7  
Percent similarity: 67.665 Percent identity: 44.611

alignment\_block:  
US-09-769-787-162 x AF307053 ..  
Align seg 1/1 to: AF307053 from: 1 to: 17854  
15 AspvaiProvaliIleglUsnleuAaSnIleThIleProlysgIysarie 31  
||||:||||| :|||:||||| :|||:|||||  
14982 GATATTCCCTCCTAGAG.....GCAAAACCGGAGACGTT 15016

```

31 uPheThrLeuGluGlyAlaSerGlyCysGlyThrThrLeuLeuArgm 48
   ::::::::::::::::::::|::::::::::::::::::|
15017 TTTAACTTCTCGCGCCGAGTGGCTGTGGAAGAACACACACTTAGAA 15066
   ::::::::::::::::::::|::::::::::::::::::|
48 ecLleAlaGlyPheAsnSerIleGluGlyGlyLeuPheTyrPheAsp 64
   ::::::::::::::::::::|::::::::::::::::::|
15067 TTAATAGCTGGTTTGAAGCCAGCAAGAGAGATATATCTTTATGAT 15116
   ::::::::::::::::::::|::::::::::::::::::|
65 ThrLysIleAsnAsnMetGluProSerLysArgAsnIleGlyMetAla 81
   ::::::::::::::::::::|::::::::::::::::::|
15117 ACCGTTATGAACGATGTTCTCCATATGAAGAACATAGGATGCTCTT 15166
   ::::::::::::::::::::|::::::::::::::::::|
81 ecLlnAsnTyrAlaIlePheProHisLeuThrValArgAspAsnValAla 98
   ::::::::::::::::::::|::::::::::::::::::|
15167 TCAGATTTATGCTCTCTTCCCATATGACGGTTTATGACAACTCTCCT 15216
   ::::::::::::::::::::|::::::::::::::::::|
98 heGlyLeuMetGlnLysLysValProLysGluGluLeuIleGlnGlnThr 114
   ::::::::::::::::::::|::::::::::::::::::|
15217 TCGGTTTAAAGCTCAGAAAGCTCTCTAAGAGACATTAAGACGAGTC 15266
   ::::::::::::::::::::|::::::::::::::::::|
115 AsnLysTyrLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysPr 131
   ::::::::::::::::::::|::::::::::::::::::|
15267 TCATGGCTCTTGAGCTGTGGGTTGAAGGATTTGAAGACCGTTACCC 15316
   ::::::::::::::::::::|::::::::::::::::::|
131 oAspLysLeuSerGlyGlyGlnGlnGlnArgValThrLeuAlaCysAlaL 148
   ::::::::::::::::::::|::::::::::::::::::|
15317 CGAGCAGCTGACGCGCGGCTCAGCAGAGAGAGTGCACCTGGCAAGGCTT 15366
   ::::::::::::::::::::|::::::::::::::::::|
148 euAlaValAsnProSerValLeuLeuMetAspGluProLeuSerAsnLeu 164
   ::::::::::::::::::::|::::::::::::::::::|
15367 TGGTGATTAACACCCAGCTTTGCTTTAGATGAACCCCTTCCATCTTC 15416
   ::::::::::::::::::::|::::::::::::::::::|
165 GluAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnH 181
   ::::::::::::::::::::|::::::::::::::::::|
15417 GATGCAAAATAGAGAAAGAGCTTAGGGAGAAATTAAGATCCAAAA 15466
   ::::::::::::::::::::|::::::::::::::::::|
181 seLlValGlyIleThrThrValTyrValThrHisAspGlnGlnGluAla 198
   ::::::::::::::::::::|::::::::::::::::::|
15467 AAGAGCTGGAATACACTACATCTACCTACGCACTGTCAGAGAAAGCTA 15516
   ::::::::::::::::::::|::::::::::::::::::|
198 ecAlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGlnGln 214
   ::::::::::::::::::::|::::::::::::::::::|
15517 TGGCAATAGTATAGATAGCGGTTATGAGCGTTGGAAGATAGAGCAG 15566
   ::::::::::::::::::::|::::::::::::::::::|
215 IleGlyArgProLysGluLeuTyrHisLysProAlaAsnGluPheValAl 231
   ::::::::::::::::::::|::::::::::::::::::|
15567 ATTGGGAATCCATGGATCTCTACTACCAACCCAAAAAATGAGTTGTAGC 15616
   ::::::::::::::::::::|::::::::::::::::::|
231 aThrPheIleGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgS 248
   ::::::::::::::::::::|::::::::::::::::::|
15617 CAGGTTTATAGCTCTTCTAATCTTTAGAGGTTAAACGACAGAGAGGAA 15666
   ::::::::::::::::::::|::::::::::::::::::|
248 erAspGlyAlaTyrIle.....ValPheSerAspGly 258
   ::::::::::::::::::::|::::::::::::::::::|
15667 AA.....GCATATGTTGGAAACCTTGTCTTGAAGTTAAACGGAAGG 15710
   ::::::::::::::::::::|::::::::::::::::::|
259 TyrAlaLeuArgMetProAlaLeuAspGlnValGluGlnAlaIleH 275
   ::::::::::::::::::::|::::::::::::::::::|
15711 .....AGAGTAA 15718
   ::::::::::::::::::::|::::::::::::::::::|
275 sValSerIleArgProGluGluPheIleLysAspGluSerGlyAspIleG 292
   ::::::::::::::::::::|::::::::::::::::::|
15719 GATATTTCTTAGGCGCGAGAGC...GTTATATCGAAGAGGAGAGC.... 15761
   ::::::::::::::::::::|::::::::::::::::::|
292 LuGlyThrIleArgAspSerValTyrLeuGlyLeuAsnThrAspTyrPhe 308
   ::::::::::::::::::::|::::::::::::::::::|
15762 .....ATGGAGAAATAGTGCACCTACGAA 15785
   ::::::::::::::::::::|::::::::::::::::::|
309 IleGluThrGlyPheAlaSerLysIleGlnValSerGluGluSerThrPh 325
   ::::::::::::::::::::|::::::::::::::::::|
15786 GTGCTCCCGGT..... 15797

```

```

325 ecLlGluAspLeuGlnLysGlyAsnArgIleArgLeuArgIleAsnThrG 342
   ::::::::::::::::::::|::::::::::::::::::|
15798 .....AGATACGGCTCAAGATTGATGTGG 15822
   ::::::::::::::::::::|::::::::::::::::::|
342 In 342
   ::::::::::::::::::::|::::::::::::::::::|
15823 AG 15824
   ::::::::::::::::::::|::::::::::::::::::|
seq_name: gb_ba:SME591790
seq_documentation_block:
LOCUS SME591790 323450 bp DNA linear BCT 16-AUG-2001
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.
ACCESSION AL591790 AL591668
VERSION AL591790.1 GI:15075230
KEYWORDS
SOURCE
ORGANISM
Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
1 (bases 1 to 323450)
REFERENCE
AUTHORS Capela,D., Barloy-Hubler,F., Guzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetalle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
From the Cover: Analysis of the chromosome sequence of the legume
symbiont Sinorhizobium meliloti strain 1021
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9877-9882 (2001)
11481430
PUBMED 2 (bases 1 to 323450)
Guzy,J.
REFERENCE
AUTHORS Direct Submission
Submitted (26-JUL-2001) Guzy J., Submitted on behalf of the MELILO
EU Consortium
COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Génétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, unite de Biochimie physiologique,
Universite catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Guzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
location/Qualifiers
1. 323450
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
155. 1519
/gene="SMC01575"
155. 1519
/gene="SMC01575"
/function="miscellaneous: hypothetical/partial homology"
/feature="product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=1
/evidence="not experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC46787.1"
/db_xref="GI:15075231"
/translation="MTPATRTGSLFPAALAMELAPANAQDAYGGYKGGDVMYLT
PEPRDRIYGEAEVHAMRDRGRRTVVDPMGNIVATVPNDGVRGQRREYGRDGY
PEPRDRGYGYSERPEFGALPEYRDIAPEVPEREDLPNSLPISLSDREAAVDPQYDPP

```

LAQPMPPAMVTYTKSRRAETALQVFLDREGSPGVIDGKMSNTKAIEMOQATGET  
LDPNNTEDILERLFRNGGLPITTYTITADAGAPVASIPEDYARAKOLPHISFTSVT  
EMLGEKFMHDEARLYLRELNPVDYSIPGTTIKVYNGPKKGVARIVAKKQVLAIV  
DEAGKLIAAYPSTIGSDTSPSGTGVHERIAIFDGYYNPKINFOQGANDEILLOLP  
GPNPGVTGVMIALSKPTYGIGHETPEPSKIGKQSGCVRLTWMDATELGKMWSTGV  
FRVD"

gene  
1707..2996  
/gene="SMC01576"  
1707..2996  
/gene="SMC01576"  
/function="small molecule metabolism"  
/note="product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence-not\_experimental  
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"  
/protein\_id="CAC46788.1"  
/db\_xref="GI:15075232"  
/translation="MATERALPMLHATAPAPRTAPLAGDLTVETAVYGGGTGLSA  
ALHAEFTGIRTAIVETARIMGEGSGRNVGLVAGWQNPDDLIATLKGAAGRLIDEL  
GDGSEFVYDLVAKKHICEEAVANGTLHNSVAGEGLKEIREPAQKKGAPVEVLSAE  
KAHALSGAGEFTGALLDRRAGTIPLAYARGLARALAGAEIFDTPLLAASROGL  
WNLKPGTAVTARHVIATNAGSLVTVGPMKRYOEELTILPYFOFATNPDPNAA  
ILPERGAMDTGLVMTSEFMDQNRILFGISRLDAIAGTHRAFAARSVRKLFPYIG  
DERFEHMDGRIGMTNNLPAHVLAPNVVISGNGVAGTVEFGALAHVHVGDT  
SAIPLAETPVTPDPRTLSAFAVHGAQAKHIDKRF"

gene  
complement(3024..3740)  
/gene="SMC01579"  
complement(3024..3740)  
/gene="SMC01579"  
/function="miscellaneous: hypothetical/global homology"  
/note="product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence-not\_experimental  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAC46789.1"  
/db\_xref="GI:15075233"  
/translation="MAVITFPAKAGGAGTAAALITSTELAROGNRVVLDADPORMI  
TMSHSGSRVANLEVISHTVPASLPCHIRELKGADPIYIDLAKAKDAIVLALGSLD  
HYLIVOCGAMDAKGAVOILEIRIGEKARIRINHSVYLVTVNSLVTTTRALQTTKAL  
LASRGVSALDTPIVERVAIRETEFCGGTLLQMDPNRVSINLDAKRENAVALAEVONL  
PYTARRALMSRISALPRAA"

gene  
complement(3949..5094)  
/gene="SMC01582"  
complement(3949..5094)  
/gene="SMC01582"  
/EC\_number="1.1.1.-"  
/function="small molecule metabolism"  
/note="product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence-not\_experimental  
/product="PUTATIVE ALCOHOL DEHYDROGENASE PROTEIN"  
/protein\_id="CAC46790.1"  
/db\_xref="GI:15075234"  
/translation="MTITANNSYPAVKEGAGRIKELADHCKALGIRKPLVYDRLGLA  
PMATQOALDILEAGLRAIFADVDPNPBNRNLGAGKAFAPDGGHDGVAFGGSGGL  
DLGCVAFMAGOTRPVMDPEIDGMMWTASVEGLAIPYAVPTTACTGSEVGRASITN  
SASHVKVTFHFKPLGVITCDPELVGPVYITTAGTMDAFAHGLEAVSFFHPMS  
AGIALEGMRLVKEYLPRAVKGADLEARANNMSAAGVAVFOKGLGAIHLSHPVGA

LYNTHHGKTNVAVMPVLRNRSATIEKIGRAAVLGIAGFDEGTYDVYVLRBELGV  
PDKLSALGVGDRIDEMAKEMAIVPTFAGNVEVETLDAERKLFACETI"

gene  
complement(5091..6476)  
/gene="SMC01588"  
complement(5091..6476)  
/gene="SMC01588"  
/EC\_number="1.2.1.-"  
/function="small molecule metabolism"  
/note="product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence-not\_experimental  
/product="PUTATIVE ALDEHYDE DEHYDROGENASE PROTEIN"  
/protein\_id="CAC46791.1"  
/db\_xref="GI:15075235"  
/translation="WTMTRCVSPVDGEVYAEPAIPLEMAHQVAHAHARLAKANARRP  
LDERVKVLAVAGVARNEMVDEVPDLAQMKRPVAYGGEFGRFENRSYVASTADAL  
KPLVVEESDREPERIAREPHGVVFIAMNYPYMTAINTVAPALMAGNTVILKNASQT  
ILVGERVRAFTIAGVPADVFOFONLELDHDTTAAIIAASKFDFINFTGSEVGRSIERA  
AAGTPTGIGLELGGKDPGVYMEDADLDAADTLMDGATVNSGCCCGGTERIYVESLY  
DAFERKSAVWSNRYKLRPLDPEPTTLGPMANKRRATYRNOYADAISGARAILDPKL  
FPQDDGAVYLAPOVLYVDHSMEEFRETTECPAVGIMVKSDAEAIELMDSRIGLTA  
SLWTKDADARARIGDLDETGTVMNRADYLDPLALCMVGKVKETGRGSLSYLGFHNLTR  
PKSYLLKRVTA"

gene  
complement(6674..8038)  
/gene="SMC01594"  
complement(6674..8038)  
/gene="SMC01594"  
/function="miscellaneous: hypothetical/global homology"  
/note="product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/transl\_table=11  
/evidence-not\_experimental  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAC46792.1"  
/db\_xref="GI:15075236"  
/translation="MSYFEEELKEDVAAGRIDTVLACVDMDQGRIMGRKRRFAEVES  
AMKETHSCNYLLADMEVETPGYKATSMERGYSQDYTKPDLSTLRIRPMDEGTAIVL  
CDMLDHDHAEVHPSPRAILKQVYARLEAMGFAYMASELEFLEFDSYDARLSGYR  
DLQASGVNEDYHIFOTPKEDVMAIRINGOGAIPVENSKEASAGOEELNRYAD  
ALTMAADRAIITKNGKETANORGRAITFLAMNNSAASSSHIHOSIMS KCGENPLRF  
DKNGOYGSSELMRHHYVAGQLAHASVYVFLAIVYINSTRFPAAGTFAPKATWSKDNRT  
AGYRIGCGESKAIRIECRVGGSDLPNPLAFAALIAAGTAGIENKMELEAPVGDAYOG  
KEVREIPHTLREAGALSGSKMLRAAFEEVEVDHYVHAWEEOOEYDRRVYDWEVARG  
FERA"

gene  
complement(8131..9678)  
/gene="SMC01597"  
complement(8131..9678)  
/gene="SMC01597"  
/function="cell processes: transport of small molecules;  
amino acids, amines, peptides"  
/note="product confidence : putative  
Gene name confidence : hypothetical  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1

alignment\_scores:  
Quality: 678.50 Length: 377  
Ratio: 2.590 Gaps: 5  
Percent Similarity: 69.496 Percent Identity: 39.523

alignment\_block:  
US-09-769-787-162 x SME591790



Align seg 1/1 to: SME591790 from: 1 to: 323450

```

1 MetSerGluLeuLeuIleAsnAlaLysLysIleThrHisValPr 17
14647 ATGGCATTCCTGCATATTCACCAATATCTTCGGCCCGGTCA 14696
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
14697 GGTGCGCATATTCGACATGGGGATCGACAGGGCGAGTGTCTCT 14746
34 euleuGluAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
14747 TTCTCGGCCCTCCGGCTGGCGACAGCATCTCGGCATGATGCC 14796
51 GlyPheAsnSerIleGluGlyGlyPheThrPheAspThrLysII 67
14797 GGTTCGAGACCCGTCGGCGCTCATCTTCATCGACGGCAAGCA 14846
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGluAsn 84
14847 GGTGGCGCTGAGCCGACACGACATATCGCATGTGTTCAGGCGCT 14896
84 yAlaIlePheProHisLeuThrValArgAspAsnAlaPheGlyLeu 100
14897 ATGCGCTCTTCCCAATGATGTCACAGCAATGTCGGCTCGGCTC 14946
101 MetGluLysValProLysGluLeuIleGluGlnThrAsnLysTy 117
14947 AAGTCGCCGCGCTCCAGACGAGATCGAGCGGCTTAAGCAT 14996
117 rIeuGluLeuMetGlnIleAlaGlnTyAlaAspArgLysProAspLys 134
14997 GCTCGACATGATCAAGCTAAGCATCTGGCGGACCGCTTCCCTATCAAT 15046
134 euSerGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
15047 TGTCCGGCGGACAGACAGCGGTGGCTTCCGCGCCCTCGCGCTC 15096
151 AspProSerValLeuLeuMetAspGluProLeuSerAsnLeuAlaTy 167
15097 AACCCGACGCTGCTCTCCGACGAGCCGCTCTCCGCGCTCATGCA 15146
167 sIeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184
15147 GATCCGCATCTCCCTCGGAGAGATCCCGCATCTCAGCAGCATCG 15196
184 yIleThrThrValTyValThrHisAspGlnGluAlaMetAlaIle 200
15197 GCATCACACAGGTATTCGTAGCAGCAGACAGAGAGGCACGTGCATC 15246
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
15247 TCCGACCGGATCGTCGTCAATTCGCGCGCGCATCAGATCGGCAC 15296
217 gProLysGluLeuThrHisLysProAlaAsnGluPheValAlaThrPheI 234
15297 CCTTTCGAGATCTACACACCCCGGACAGCTTGTCCTCTCTCG 15346
234 leGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
15347 TCGGCACGCTCAACATCATCGAGGCAAGGTGGCCGATCCCGAAGGCG 15396
251 AlATyTlIleValPheSerAspGlyTyAlaLeuArgMetProAlaLeuAs 267
15397 GCGATGACCATCGCGGCGCCCATCTCAGTAAGAGCGATTGCGCG 15446
267 gAlnValGluGluAlaIleHisValSerIleArgProGlu.....G 282
15447 GCGAAAGCGCGGACAGCATCTCGCTGCGCTAGCGCGAAGCAGCGCT 15496
282 luhelIeLysAspGluSerGlyAsp.....IleGluGlyThrIleArg 296
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

15497 CGATCGCCGAGGGCGCCAGGGCGGATACCGCATCTCCGGGCGAATGCTA 15546
297 AspSerValTyIreGluY..... 302
15547 TCGAGGAGCTTTCGCTCCGTCATCCGACCAACCTCGGGTCGGCAA 15596
303 .....LeuAsnThrAspTyPheIleGluThrGlyPheAlaSerLysI 317
15597 AGCATCATCTCTTCGATATGTTCACAGCATCGGGCATGACCCACCG 15646
317 leGlnValSerGluGluSerThrPheGluGluAspLeuGlnTyGlyAsn 333
15647 TC.....GCCGCGAG 15657
334 ArgIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAla.. 349
15658 AGGTCACGCTGCGTTTCGCGCAAGACCTCGTCATCGCGAATA 15707
350 .AspGlySerGluAsnLeuIleLysGlyVal 359
15708 GACGGCGCGAATGAAATTTTCAGGGCGGTG 15738

seq_name: gb_ba:AE008131
seq documentation block:
LOCUS AE008131 11311 bp DNA linear BCT 18-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 189
of 254 of the complete sequence.
ACCESSION AE008131 AE007869
VERSION AE008131.1 GI:15157288
KEYWORDS
SOURCE
ORGANISM
Agrobacterium tumefaciens str. C58 (Cereon).
Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE
1 (bases 1 to 11311)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete genome sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
JOURNAL
2 (bases 1 to 11311)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
AUTHORS
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
JOURNAL
FEATURES
location/Qualifiers
1..11311
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/strain="C58"
/db_xref="taxon:181661"
77..1072
/gene="AGR_C_3892"
77..1072
/gene="AGR_C_3892"
/note="probable permease of ABC transporter PA3252
(imported) - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=1
/product="AGR_C_3892p"
/protein_id="AAK87895.1"
/db_xref="GI:15157289"
/translation="MKRTKIRADSVYLTGCARBPLETDRFNGYWGCGMPNNNG
GPSDAGRKLPLNMIGVPPFAVFMFLILPMPKIVIGAFQTDGSDFTDNTAGLFTSS
ILAAVWISIKISLASAALGCLIGFAVAANAVALGILPQIRGPLTFSGVASQFAGVPL
AFATATGIPVGLTFLPKQIGIDLRILGNIISFPGILVTYLFPIPLMILITPA
LDGLKREMRBAEILIGATGLOYRMVAAPILPILGMSLLFANAGAVATAATG
SSINIVPLTFAQIRGDVLTGSRPHLGALAFGMIVLTGIANNLYIMLARSRWLK"
1060..1851
/gene="AGR_C_3893"
1060..1851
/gene="AGR_C_3893"
/note="probable ABC transport system permease Cj0731

```

```

(imported) - Campylobacter jejuni (strain NCTC 11168)
/codon_start=1
/transl_table=11
/product="AGR_C_3893p"
/protein_id="AAK87896.1"
/db_xref="GI:15157290"
/translation="MAEMKRFAMGALITGGLYLFALRPLGTMFTSLKMRGEVSDAY
GAVLDPRDFQEFSSVVALFFIVEGLLVLPFAVWRKLDGLRPIEFITLLPLV
IPAVIVEQYIKLVNTSSMLPLGTTFGNLLMFGVATLALMYRAVDGTGLTIDV
SUTLEAOSIGAGWTLILSRILIPNLVAVLSCGAFLEFAIVIEFTMAALLNPAECP
YMQLLGANRYEPAAALAVYISGITWCGLGILQVSRXQKAPKA"
1865..2923
/gene="AGR_C_3896p"
1865..2923
/gene="AGR_C_3896p"
/feature="ABC transporter, ATP-binding protein VCA0602"
(imported) - Vibrio cholerae (group O1 strain N16961)
/codon_start=1
/transl_table=11
/product="AGR_C_3896p"
/protein_id="AAK87897.1"
/db_xref="GI:15157921"
/translation="MSFLITSNKTSFSGSYOVYVHDNMAIEKEEYVSLPSGCGKTT
VLMRAGEPIPGSSVINGKQITDLPDRNINMGFOAYALPNNMYVNNVAGKLV
AGKPRAEIDARKEMQILHLEHADRYPYKMSGGQOORALAKALAKREYLLDPR
LSADLAKVSRLEELRIQDQKIGITTFVTHQEEALSISDRIYVNNNGRADIGSP
FDIYKPAIRFAVSEFGLNLLIDATVDDSSNTRIYGEQGITLPKPYDANGKEITLTA
LRPEGSIGSGDKDPAVILGLVWSSQFLSVIRTLDLGSSITSPDFMNPPTAPPAV
GERVTLKFAAGSLMIKD"
2988..4061
/feature="AGR_C_3897p"
2988..4061
/feature="AGR_C_3897p"
/feature="ALCOHOL DEHYDROGENASE (ADH-HT)"
/codon_start=1
/transl_table=11
/product="AGR_C_3897p"
/protein_id="AAK87898.1"
/db_xref="GI:15157922"
/translation="NMLPRNDENMRAIYERFGEIVSAGRNVTREFCTGRVYIYAAK
TGICSDMHGMCHGMDIFLPHYRGHEFAGVISAAGRNTYRFTGDRVYIPYSCGH
CHEGSGNQOVCETQPOPFTHGSRREYVAIDYADQNLVHLDESMYATAAGLCFF
ATSFPAVYIDQGLKGGEVALHGGCGGVASLIMIGAGLQVVAIDAEKLEIARQL
GATPAINSVSDVAEAVNDIGCGGHVSDALDIGHQTCNSINLRRRRHVOVGMGL
LADHAPAIAPMARVLAHLEIYISGHQMAMRYEDMLAMIESGLAPKELIGHISTEE
LAVALPGMDRPFESGSIIDREF"
4408..4730
/feature="AGR_C_3900p"
4408..4730
/feature="AGR_C_3900p"
/feature="hypotheetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_3900p"
/protein_id="AAK87899.1"
/db_xref="GI:15157923"
/translation="MSYATLFSAGALFSYIFASVAAERKDRLEPAISPGSGEETALY
DLPNPAFLAEONTDIDGLINDSGGVLYLVGDRYTRISDAVDGRKSLGDPDTQ
HRLERAAERKQKERRNLATFFLYLVYGFVLVYLKGLDAVRMFRLLTMVDTBKQAA
EPHDLAKKKRLVRRQYSRPRDTSVAASPAPVPEASPTGGTARAFGRRTK"
5063..5839
/feature="AGR_C_3902p"
5063..5839
/feature="AGR_C_3902p"
/feature="AGR_C_3902p"
/feature="AP001508) BH0315
unknown conserved protein in others"
/codon_start=1
/transl_table=11
/product="AGR_C_3902p"
/protein_id="AAK87900.1"
/db_xref="GI:15157924"
/translation="MFLDSVAMKEKPEGRREEXYPSVAPAMHLELKEFKTPIFFAGN
NGSGSTLEEGLAAGATAAIVNMHGDVKGDLVYLQHAETVAKAFYPAKKYPIKIMEFAR

```

[illegible]

gene  
/gene="Atu2148"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, membrane spanning protein"  
/protein\_id="AAU43138.1"  
/db\_xref="GI:17740613"  
/translation="MPSNNTGSGAGRLPLNMIGVPEAVFMFLIPTRKIYIG  
AFQDTGSEFLDNIAGFTSSIIAAWISIKISIASAAGCLGPAVAAVYGLDPO  
RIKGPLTFSGVSOFAGVPAFAFIATPLGVLVPLKTOIGDILRPNITISFM  
GLVYTLFQIPLMITITPDLGDKRREAREAITGATGLOTRVRAFPILPSLIG  
TMSLTPANAGAVATAIALTGSSLNITVILFQIINGDVAGNPHIGYALAFGMITLIG  
IANALYIMLARERERWLK"  
/gene="Atu2149"  
/gene="Atu2149"  
/gene="Atu2149"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, membrane spanning protein"  
/protein\_id="AAU43139.1"  
/db\_xref="GI:17740614"  
/translation="MKRFAMGALTFGLYFALPLIGMTNFSIKMRGEYSEDAVKV  
LTPDRETFSSYVMALEFIVGVLLVPTAVVWRLKLGRLPYIEFTILPLVPA  
IYIVFGYIRLYNYSWMLPTGFTGTLILMEGATLIPYMRVADTGRTIDVSTL  
TEAOSIGAGMTTILSRITLPLVAVVLSGRTFAIVGEFTMALNLRPAFGPIQO  
LIGANRAYEPALAVISFGITWGCLGILQVSRVOKGAPPKA"  
/gene="Atu2150"  
/gene="Atu2150"  
/gene="Atu2150"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, nucleotide binding/ATPase  
protein"  
/protein\_id="AAU43140.1"  
/db\_xref="GI:17740615"  
/translation="MSFLLTSLNKKSGFSGVOVVDHFNMAIEKGEVSYLPGSCGKTT  
VLRMAGFEIPTGGSIVNGKQDTTLRPQNRNGVQAVALEPNNVYENVAFGKLV  
AGKPKAEIDARVKEMLQILHLEHLDRIYQOMSGQOQVRLALAPYEVILLDEP  
LSADAKIRVSLREERIOLOKGIITFYTHQEEALSISDRIVVMNGRADQISDP  
FDIYNPATRPVASFVGTNLIDATVVDSSNTRVGEBOITLPRKPVDAANGKRTILA  
LREAGSLGDAKGVAVISGLVTSQFLGSVYTRLDLGGSTLSFDMFPDGPAPPAV  
GERVTLKFAAGDLMLIKD"  
/gene="Atu2151"  
/gene="Atu2151"  
/gene="Atu2151"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="alcohol dehydrogenase"  
/protein\_id="AAU43141.1"  
/db\_xref="GI:17740616"  
/translation="MRALYERFGEPTPVVSLPDPAPSDGVVIAVAKGCLCRSDMHG  
WMGHDDIRLPHVPGHEFAGVISAGRNTRFTGDRVTVPEVSGKHGHEGSGMQO  
VCETQPGFTHMGSPAEVYADYDONLVHLEPSVSATAGLGRFATSPFAVTDQ  
GRKGGEMLVHGGCGVLSAIGAGLAOVYADIDIAEDKLEIAROLGATATINRS  
VADVAVARDIINGCAHVSVDALGHPTQCNCSNRRGRHYQVYMLADNAPRIP  
MARVIAHELEITGSHGQAMRTEDMLAMTESGRLAPEKLIGRHISLTAFAVALPGMDR  
EQESGISIIDREF"  
/gene="Atu2152"  
/gene="Atu2152"  
/gene="Atu2152"  
/note="putative; ORF located using Glimmer"

gene  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAU43142.1"  
/db\_xref="GI:17740617"  
/translation="MSYTRLPFSAGALFSYIFASVAEARKDRLFAISEGCGETIAY  
DLPNTPAFIABEGNTLDLGLNSDSGSGVLYIGDRYTRLSDDVGLSRGLPDDTQ  
HRLERAAREKQERNMLAIEFVLVVGFFLVYKGLDAVRWRLLTMVDPDKGOAA  
EPHDPKAKRRKLVROYSRPPTDSVAASPVPPEASPTGTARAGRRRK"  
/gene="Atu2153"  
/gene="Atu2153"  
/gene="Atu2153"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, nucleotide binding/ATPase  
protein"  
/protein\_id="AAU43143.1"  
/db\_xref="GI:17740618"  
/translation="MLFDSVAMKEKPEGSEYPPSPAMRHLERLEFPTTFPAGN  
NSCKSTLLEGIAAGMTATAIGNHGDYKGLYLQHAETVAKAFYARRKYPKIRMEFLR  
ADVLGATIRQENALDDPRMERERALKGEDPPESTPEITFGYIRNNSLDRSIBCG  
FLDIMHRLHAGLYFLDEPESPLSPQKQLELAIIROAADSGGOLLITATSHPVLLAI  
PAAIYVFEDECDITERLYDELINIAFLRFRFLDRPSKYLDD"  
/gene="Atu2154"  
/gene="Atu2154"  
/gene="Atu2154"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="MFS permease[proline/detaine]"  
/protein\_id="AAU43144.1"  
/db\_xref="GI:17740619"  
/translation="MSPTTPAARPLNSODEFTLGSALGALFEYDITTFPEFATV  
IGHLEFPPPMPLWVIOFEGIFAGVLYRPLGGIVLAHYGRYRKRKRVAFSILMA  
LSTIGMALMPRTATIGVAPILLITLIRMOGASIAIGEPGATFYAEHVPERVGLAC  
GFLTSGLSRGIMUGSLINAFINSLTPTDEVACYNARIPRLLGGVGLIAYIRPLEE  
PTVYEMKRSKSLTDKPLGLVLRKHRRKQVITISALIVLWSAIVYITLMTATFLQKL  
YVYTPPQALAGTSFGLFLIFGVIIAGALIDRVGSGIFPMGASIFEGIATFTFYSYAG  
TSELEMTFVYGMVGSVAGAVAPVVMVRAFPASVSFSGISPAVVSVAVEFGGLPIG  
VTTALAINMAWYLVFAFAFPAIAGLYLHURGSEVSHVIEBLAARS"  
/gene="Atu2155"  
/gene="Atu2155"  
/gene="Atu2155"  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="methionine synthase"  
/protein\_id="AAU43145.1"  
/db\_xref="GI:17740620"  
/translation="MDDLFCEPGEKAGRDGAEIFKALRDASERILLIDGAMGTIOGL  
GPEDEHFRGDRITGCACGKGNNDLLITQPAIDIEIHRYAMACADILERTFTSSSTR

alignment\_scores:  
Quality: 671.00 Length: 331  
Ratio: 2.750 Gaps: 4  
Percent Similarity: 73.716 Percent Identity: 44.109

alignment\_block:  
US-09-769-787-162 x AE009165 ..  
Align seg 1/1 to: AE009165 from: 1 to: 12416  
1 MetcergtutlelystilleAsnAlalystlylelyrthsaasValPr 17  
||||||| ::: ||| ||||| ::: |||||







```

|||||
1 ATGAGTGAGATCAAAATTATTATACGCCAAAAAATCTACCAACGATGTCCTCC 50
17 oval11GluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
51 TGTATTGAGAAATTTGAAACATTACAAATTCAAAAGAAAGTCTCTTTACCC 100
34 euLeuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
101 TTTCTGGAGCTTCAGATGTGGAAAAAGACCTCTCTTCGTATGATTGCA 150
51 GlyPheAsnSerIleGlyGlyGlyLysPheAspPheThrLysIle 67
151 GGTTCACACGATATCGAAGGTGGAATTTTACTTCGATGATACAAAAAT 200
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsn 84
201 CAATTAATATGAAACCCAGCAACCAATATCGGATGTTTCCAAAACCT 250
84 yAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
251 ACGTATTTTCCACATTTGACTGCCAGAACGTTGGCTTTGGTCTT 300
101 MetGlnLysLysValProLysGlyGlyLeuIleGlnGlnThrAsnLys 117
301 ATGCAAAAGAGGTTCCAAAAGAAATGATTCACAGACCAACAAAGTA 350
117 fLeuGlyLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLys 134
351 TCTTGAACCTCATGCNAATGCTCAATATCGGATCGAAAGCCCATTAAC 400
134 euSerGlyGlyGlnGlnAlaArgValThrLeuAlaCysAlaLeuAlaVal 150
401 TCACTGGTGACACAAACAGCTGCACCTGGCATCGCGCTTAGCGGTT 450
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLys 167
451 AATCCAAGTGTCTCTCATGAGAGACCACTAGTAATCGAGAGGCCAA 500
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
501 ACTTCGGCTGGATATGCGTCAAGCATCCGAAATCCAAACACCAAGTGG 550
184 LylIleThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200
551 GAATTACAACACTGTTATGTAAACCCACAGCAAGAAAGCATGGCTATT 600
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217
601 TCAGACCAAAATGCTGTTATGAAAGATGGGTGATCCAAACAAATCGGCG 650
217 gProLysGlyLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
651 ACCAAAAGAACTCTATCATTAACCAAGCTAATGAGTTTGTGCAACCTTTA 700
234 leGlyArGthrAsnIleIleProAlaAsnLeuGlyLysArgSerAspGly 250
701 TCGAGCCCAAAATATATTCCTCCAACTCTTGA AAAAGAGGAGCGCG 750
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
751 GCTTATATCGCTTTTTCAGATGCTATGCGCTTGAATGCGAGCTCTTGA 800
267 gAlaValGluGluGlnAlaIleHisValSerIleArgProGluGluPhe 284
801 TCAGGTTGAGAGGACCAACTATTCTGTAAAGCATTCGTCGCCAAGTTTA 850
284 leLysAspGluSerGlyAspIleGlyLysThrIleArgAspSerValTyr 300
851 TCAAAGATGATCTGAGATATTGAAGAACTATTAGAGATAGCGCTTAT 900
301 leuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317
|||||

```

```

901 CTTGAGCTAAATACGATTAATTCATTGAGACAGGTTTCCCTCAAAAAT 950
317 eGlnValSerGluGluSerThrPheGlnGluAspLeuGlnLysGlyAsnA 334
951 TCAAGTTAGTGAAGATCACTTTTGAAGAAGATCTCAAAAAGGCATC 1000
334 rGlyIleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAsp 350
1001 GTATTCGCTTACGAAATCAATACGCAAAAATTTAAACATCTTTTCGCAGAT 1050
351 GlySerGlnAsnLeuIleLysGlyValAsnHisGlyThr 363
1051 GGTTCACAAAACCTGATTAAGAGCTACACCATGCAACG 1089
seq_name: /SIDSI/gc9data/geneseq/geneseqn-emb1/NA1998.DAT:AAV52260
seq_documentation_block:
ID AAV52260 standard; DNA; 9578 BP.
XX
AC AAV52260;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae; genome fragment SEQ ID NO:127.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN W09818931-A2.
XX
PD 07-MAY-1998.
XX
PE 30-OCT-1997; 97WO-US19588.
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences; - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 876-882; 1409pp; English.
XX
PS
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and

```



CC pharmaceutical compositions and vaccines for *S. pneumoniae*.  
 xx Sequence 9578 BP; 2886 A; 1653 C; 2116 G; 2923 T; 0 other;  
 SQ

## alignment\_scores:

Quality: 1843.00 Length: 363  
 Ratio: 5.077 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-769-787-162 x AAV52260/rev ..

Align seg 1/1 to reverse of: AAV52260 from: 1 to: 9578

```

1 MetSerGluIleLysIleIleAsnAlaLysIleTyrHisAspValP 17
|||||
8112 ATGAGTGAATTCATAACATTACATTCCTCAAGAGAGAGTCTTACCC 8063
17 oValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
|||||
8062 TGTATTGAGATTTGACATTACATTCCTCAAGAGAGAGTCTTACCC 8013
34 euLeuGlyIleSerGlyCysGlyLysThrThrLeuAsnArgMetIleAla 50
|||||
8012 TTCTTGAGCTTCAGATGTGGGAAACGACCTTCTGCTATGATTGCA 7963
51 GlyPheAsnSerIleGluGlyGlyPheTyrPheAspAspThrLysI 67
|||||
7962 GGTTCACACAGTATCGAAGTGGAGATTTTACTGATGATCAAAAAAT 7913
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGluAsn 84
|||||
7912 CATTAATATGAGACCAACGCAATATCGGATGCTTTTCCAAACT 7863
84 yAlaIlePheProIleuThrValArgAspAsnAlaPheGlyLeu 100
|||||
7862 ACGCTATTTTCCACATTTGACTGTCGAGACACGTTGCTTTGGCTT 7813
101 MetGluLysLysValProLysGluGluLeuIleGlnGlnThrAsnLys 117
|||||
7812 ATCCAAAGAGAGGTTCCAAAGAGAGATTCATCAACAGCCAAAGTA 7763
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLys 134
|||||
7762 TCTTGACATCATCAATGCTCATATGCGGATCGAAGCCCGATTAAAC 7713
134 euSerGlyGlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
|||||
7712 TCAGTGTGGACAACAACAGTGTACCTGGCATGCGCTTAGCGGTT 7663
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuAlaLys 167
|||||
7662 AATCCAAAGTGTCTCTCATGAGAGCCACTAGTAATCTGGAGGCCAA 7613
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184
|||||
7612 ACTTCGCTTGATGCGTCACGCCATCCGAGAAATCCACAGAGAGTGG 7563
184 yIleThrThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200
|||||
7562 GAATTACAACCTGTTATGTAAACCCAGCAGCAAGAGAGCAGGATAT 7513
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGly 217
|||||
7512 TCAGACCAAAATTCGTGTATGAAGAATGGGTATCCAAATAATCGGCC 7463
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||
7462 ACCAAAGAACTCTATCATTAACAGCTAATGAGTTTGTCGAACCTTTA 7413
234 lGlyArgThrAsnIleIleProAlaAsnLeuLysArgSerAspGly 250
|||||

```

```

7412 TCGAGCACAATAATATTCTCTGCCAATCTTGAAAAAGAGCGAGCGC 7363
251 AlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeuAs 267
|||||
7362 GCTTATATCGTCTTTCAGATGGCTATGCGCTTCGATCGACGCTTGA 7313
267 pGluValIleGluGlnAlaIleHisValSerIleArgProGluGluPhe 284
|||||
7312 TCAGTGTGAGAGACCAAGCTATTTCATGTAGCATTCCTCCGAAAGATT 7263
284 lLysAspGluSerGlyAspIleGluGlyThrIleArgAspSerValTyr 300
|||||
7262 TCAAAGATGAATCTCGAGATATTGAAGCAACTTATGAGATAGCGTCTAT 7213
301 LeuGlyLeuAsnThrAspTyrPheIleGluThrGlyPheAlaSerLysI 317
|||||
7212 CTGGAGCTAAATACGAGATTTTCATTGACAGCGGTTTGCCCAAAAT 7163
317 eGluValSerGluGluSerThrPheGluGluAspLeuGlnLysGlyAsn 334
|||||
7162 TCAGTGTAGTGAAGAAATCAACTTTGAGAAGATCTACAAAAAGCAATC 7113
334 rGluLeuArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAs 350
|||||
7112 GTATTCGCTACGATCAATACGCAAAATAATTAACATCTTTCTGCAGAT 7063
351 GlySerGluAsnLeuIleLysGlyValAsnHisGlyThr 363
|||||
7062 GGTTCCTCAAAACCTGATTAAGAGATCAACCAAGGAGACG 7024
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AAx30946
seq_documentation_block:
ID AAX30946 standard; DNA: 693 BP.
xx
AC AAX30946;
xx
DE 20-MAY-1999 (first entry)
xx
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:223.
xx
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;
xx Streptococcal infection; pneumococcal; ss.
xx
OS Streptococcus pneumoniae.
xx
PN WO9737026-A1.
xx
PD 09-OCT-1997.
xx
PP 01-APR-1997; 97WO-US05306.
xx
PR 22-AUG-1996; 96US-0025788.
xx PR 02-APR-1996; 96US-0014690.
xx
PA (SMTK ) SMITHKLINE BEECHAM CORP.
(SMTK ) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
DR WPI: 1997-503111/46.
DR P-PSDB; AAY11367.
xx
PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in
xx vaccines, drug screening, etc
xx
PS Claim 5; Page 185; 354pp; English.
xx
CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,
CC isolated from Streptococcus pneumoniae, can be used in vaccines against

```

CC streptococcal infections and in assays for identifying compounds that  
 CC inhibit or activate the activity of the proteins. The antagonists can  
 CC be used to treat an individual having need to inhibit a bacterial  
 CC protein. Vectors expressing the proteins can be used to induce a  
 CC protective immune response in mammals.

XX Sequence 693 BP; 218 A; 155 C; 145 G; 171 T; 4 other;

# alignment\_scores:

Quality: 1003.00 Length: 231  
 Ratio: 4.580 Gaps: 5  
 Percent Similarity: 94.805 Percent Identity: 90.909

# alignment\_block:

US-09-769-787-162 x AAX30946 ..

Align seg 1/1 to: AAX30946 from: 1 to: 693

```

38 SerGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsnSe 54
|||||
3 TCAGGATGTGGGAAACGACCCCTTCTGATGATTCAGAGTTTCACACAG 52
|||||
54 rllieguglyglyluphetyrPheaspapThrLysIleAsnAsnMetG 71
|||||
53 TATCAAGATGGAGAAATTTACTTCGATGATACAAAATCAATATATATG 102
|||||
71 luproSerLysArgAsnIleGlyMetValPheGlnAsnTYrAlaIlePhe 87
|||||
103 ACCCGACCAACCGCAATATCGGCTGTTTCCAAACTACGCTATTTC 152
|||||
88 ProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLysly 104
|||||
153 CCACATTGTGACTGTCCGACGACAGCTTGGCTTGTTCGCAAAAGAA 202
|||||
104 sValProLysGlyGluLeuIleGlnIlnThrAsnLysTYrLeuGluLeu 121
|||||
203 GGTTCGCAAGAGAAATGATTCACACAGACCAACAGTATCTTGACCTCA 252
|||||
121 etGlnIleAlaGlnTYrAlaAspArgLysProAspLysLeuSerGlyGly 137
|||||
253 TCGCAATGCTGCATATCGCGATCGAAGCCCGATTAACCTACAGTGTGGA 302
|||||
138 GlnGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSerVa 154
|||||
303 CACCAACAGAGTGCACCTTGCGATGCGTCTAGCGGTATTCACAATGT 352
|||||
154 IleuLeu.MetaspGlu.ProLeuSerAsnLeuGluAlaLysLeuArgLe 170
|||||
353 TCTCTCTCCATGAGGAGGAGCCACTAGTAATCTGGAGGCCAAACTTCGCTT 402
|||||
170 uasPmetArgGlnAla...IleArgGluIleGln...IleGluValAlaGly 185
|||||
403 GGATATGCGTTCAAGCCCATCCGAGAAATCCACACAGAGTTGGGGAA 452
|||||
185 lenthThrValTYrValThrHisAspGlnGluAlaMetAlaIleSer 201
|||||
453 TTAACAAGTGTATGTAACCCAGACCAAGAGAGGACCATGGCATTTTCA 502
|||||
202 AspGlnIleAla.ValMetLysaspGlyValIleGlnGlnIleGlyArgp 218
|||||
503 GACCAAAATTCCTGTATGAAAGATGGGTGATCCAAACAATCGGCCGAC 552
|||||
218 rOlysgGluLeuTYrHisLysProAlaAsnGluPheValAlaThrPheIle 234
|||||
553 CAAAGAGACTGTATCATAAACAGCTATATGATGAGCAACCTTTATTC 602
|||||
235 GlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSeraspGlyAl 251
|||||
603 GAGAGCGCAATATATTATCCCTGCAATCTTGAAAGAGGAGGACGCGCGN 652
|||||
251 aTYrIleValIlePheSeraspGlyTYrAlaLeuArgMet 263
|||||

```

653 TTAATCTGCTTNTTCAGATGAGANANGCCCTTCGAATG 689

seq\_name: /SID51/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF28538

seq\_documentation\_block:

ID AAF28538 standard; DNA: 31147 BP.

XX AAF28538;

DT 04-APR-2001 (first entry)

DE Genomic fragment #25.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;

KW bronchopulmonary; endocarditis; meningitis; ss.

OS Moraxella catarrhalis.

PN W020078968-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000MO-US16649.

PR 18-JUN-1999; 990US-0140121.

PA (INCY-) INCYTE GENOMICS INC.

PI Lagace RE, Patterson C, Berg KL;

DR WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic  
 PT compositions, and for identifying virulence factors, regulatory  
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
 PT acids -

PS Claim 1; Page 207-215; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.

XX Sequence 31147 BP; 8961 A; 6045 C; 6778 G; 9363 T; 0 other;

# alignment\_scores:

Quality: 669.00 Length: 376  
 Ratio: 2.544 Gaps: 7  
 Percent Similarity: 69.947 Percent Identity: 40.160

# alignment\_block:

US-09-769-787-162 x AAF28538 ..

Align seg 1/1 to: AAF28538 from: 1 to: 31147

```

1 MetSerGluIleLysIleIleAsnAlaLysIleTYrHisaspValPr 17
|||||
18353 ATGAGCTATATTCAATTAACCAATGCACATTAAGTCTTCGTTCAATTGAC 18402
|||||
17 ovalIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34
|||||
18403 CGTTATGATGATTGATTAATGATGAAGAGGAGATTGATTGTTAC 18452
|||||
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50
|||||

```

```
18453 TACGTGGCTTCACGCTCGGCACGTCACCTTTGTTACGTTGATTGCC 18502
51 GlyPheAsnSerIleGluGlyGlyPheTyPheAspAspThrIysI 67
18503 GGTCTGGAAACATTGATCATGAGGCTTCATATTATTAAACATCAAGATT 18552
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsp 84
18553 TACCTATTAAACCCAGAAACGCCGTATTGGCATGATTTCATAAATT 18602
84 yTrAlaIePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100
18603 ATGCTCTCTCCCAACATGACAGTGGCAGATTATGTAGATTGGATTAA 18652
101 MetGlnLysLysValProLysGluGluIleGlnGlnThrAsnLysTy 117
18653 AAGATTAAAAAGTCTCTCTTGAGAGAAAGCTGACTTAAGTCAAAACGT 18702
117 rLeuGluLeuMetGlnIleAlaGlnTyAlaAspArgLysProAspLys 134
18703 GCTTGATTGGTTGAGTTAACTTCTTTGCTCAGCAAAAACGAGATTCT 18752
134 euSerGlyGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150
18753 TATCTGGTGACAAAAACAGCGTTGCTGACCTGAGACATTAGTTATG 18802
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAla 167
18803 GAGCCGATTTGCTTTTGTAGTGGCCATTATCCGCTCTGATGCCAA 18852
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184
18853 ACTAGCTAAAGTTTACGATGCAATTTAAGCGCATTCAAAAGATTGG 18902
184 yTleThrThrValTyValThrHisAspGlnGluGluAlaMetAlaIle 200
18903 GATTACTACACTTTTGTACTCATGATGATGAGTGAACCTTTAGCGATG 18952
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGly 217
18953 TCAGACGAGAGTGTCTTCTTATTAAGGCACAGATTGACACATTCCTC 19002
217 gProLysGluLeuTyThrHisProAlaAsnGluPheValAlaThrPhe 234
19003 GCCAGACACACTATATACCTACGCTAACATAGSATTACTGCGGGTT 19052
234 lGcIyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 250
19053 TAGGACATTACAAATATA..... 19069
251 AlaTyrlIleValPheSerAspGlyTy.....AlaLeuArgMetProAl 265
19070 .....GGATATTTTGAATCAGTAAATCCAAATC 19098
265 lAleuAspGlnValGluGluGlnAlaIleHisValSerIleArgProGlu 282
19099 TGCCAAGCAACTTTCGATGATGCA.....ATAAGCCCTGAAA 19136
282 lPheIleLysAsp..GluSerGlyAspIleGluGlyThrIleArgAsp 297
19137 CATATATTGTAGATACGATGATGAGATATTCACAGGCTTATTCCTGA 19186
298 SerValTyLysGluLysAsnThrAspTyPheIleGluThrGlyPheAl 314
19187 AGAAGCTTACTGGGGAGTGGCTTACCAAGTCAGAACAAATATGCG 19236
314 aSerLysIleGlnValSerGluLysSerThrPheGlnGluLysAspLeu 331
19237 TGATATATTGATGT.....GATGTACTGAACC 19265
331 yGcIyAsnArgIleArgLeuArgIleAsn..... 340
19266 ATGGCAAAATTTAGCCAAATTAAAGTCAAATTGCAAGGTTTTTTTAATA 19315
```

```
341 .....ThrGlnLysLeuAsnIle.....PheSerAlaAspGlySe 352
19316 AAAAGATGAACCAAGTGTATCAACCTAGTACTTATTATTCTTACCTTA 19365
352 rGlnAsnLeuIleLysGlyValAsnHis 361
19366 CAAAACACTAGCTACTTAAATATGCATCAT 19393
seq_name: /SID1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:AA53495
seq_documentation_block:
ID AA53495 standard; DNA; 1146 BP.
XX
XX AC AA53495;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Haemophilus influenzae DNA for cellular proliferation protein #277.
XX
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX OS Haemophilus influenzae.
XX
XX PN WO200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US09180.
XX
XX PR 21-MAR-2000; 2000US-191078P.
XX
XX PR 23-MAY-2000; 2000US-206848P.
XX
XX PR 26-MAY-2000; 2000US-207727P.
XX
XX PR 23-OCT-2000; 2000US-242578P.
XX
XX PR 27-NOV-2000; 2000US-253625P.
XX
XX PR 22-DEC-2000; 2000US-257931P.
XX
XX PR 16-FEB-2001; 2001US-269308P.
XX
XX PA (ELIT- ) ELITRA PHARM INC.
XX
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX DR WPI: 2001-611495/70.
XX
XX PT P-PSDB; AAU35636.
XX
XX PS New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 7132; 51pp; English.
XX
XX CC The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1146 BP; 378 A; 203 C; 237 G; 328 T; 0 other;
```





PT products for the detection, diagnosis, characterisation, prevention  
 PT and therapy of infections, particularly Lyme disease

XX  
 PS Claim 1: Page 157-671, 1128pp; English.

CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs  
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.

SO Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

alignment\_scores:  
 Quality: 650.50 Length: 359  
 Ratio: 2.531 Gaps: 8  
 Percent Similarity: 71.588 Percent Identity: 38.997

alignment\_block:

US-09-769-787-162 x AAX20248 ..

Align seg 1/1 to: AAX20248 from: 1 to: 910715

6 lleleasnlaalsylsyletyrhls.....aspvalpr 17

229127 ATCTTAGAGATTAAATCTAGCTATTATTGATTAACAATGGAACAA 229176

17 ovalilegluasneuansilietrhileprolysglyserleuphetrl 34

229177 AACTTAGATTAACATAATTAAATAAATAAGAGTTTCACAC 229226

34 eueuglyalaserglycysglylysthrthleuauarqmetila 50

229227 TACTAGCCCATCGGATGTGGAAAAACAACATGATTAATAATGTGGGT 229276

51 G1PheasnserilegluglygluPhetypheaspaprtrlsil 67

229277 GCTTTTAAAGCCAAAAAATGGAAATTTATTTCTTTCTTAAGAAAT 229326

67 easnasmetgluproserlysarqasniileglymetvalpheglnasr 84

229327 ATCTAAACACAGTCCAAACAAAGAGAAATTAATACGTATTTCAAAAT 229376

84 yralalietpneprohlsleuthvalarqaspsnvalalapheglyleu 100

229377 ATGCACCTTTCCACATATGATGTGTTTGCACATATTTCATTTGGACT 229426

101 MetglnlyslsvalprolysgluGluleuileglnlnhrlnstly 117

229427 AGAATGAAAAAAGCCAAAGATATATCAAGAAAAAGAAACATC 229476

117 rleuglnleuMetglnleuileaglnlyalasparglysprosprlysl 134

229477 GCTTTCCTGATAGCAATGCCAAATAGCATACAGAAATATTACGAAC 229526

134 eusergllyglnglnlnarqvalthleualacysalaleuval 150

229527 TATTCGGGGGGGCAAAAGCAATTCATAGCAAGCAAGCAATGTGTAATG 229576

151 AspProserValleuMetaspGluProleuserasnleuglnalaly 167

229577 GAACCTAGCTTTTACTCTAGATGAACCACTTCCCGCTTGATTTGAA 229626

167 sleuargleuaspmetarqalnalalearqgluileglnhisgluval 184

229627 AATGCGACAAAGATGCAAAAGAAATTAATAAATACAGCGTCAGCTTG 229676

184 lyliethrthrvaltyrvalthrhisaspGlnuglnuialametalaly 200

||||| .....  
 229677 GAATCACATTCATATATGTACTCACGATCAGAGAGCGCATTGACATG 229726

201 SeraspGlnleuAlaValMetLysAspGluValIleGlnInleuIle 217

229727 AGTACAGAAATCGTTGTAATGAATGAAGAAATATTTCGAAATAGCAAC 229776

217 gProlysgluLeuTyrlsLysProAlaansgluPheValAlaThrPhe 234

229777 ACCTGAGAAATTTACAAATGAGCTTAACAAAGTTGTAGCGATTTTA 229826

234 leglYargthrAsnIleIleProAlaasnleuGlulysArgSerAspGly 250

229827 TTGAGAAACCAATATTTT .....GATGGA 229852

251 AlaTyrl .....IleValPheSerAspGlyTyrlAlaLeuArgme 263

229853 ACATATTAATAAAGAGCTGTTGTAGCTTCTTGTCAT .....GAAT 229896

263 tProAlaLeuAspGln .....ValGluGlnAlaIleHisValSer 278

229897 TGAATGCCCTTGACAAAGATTTGAAGCTGAAGAAGCAGTTGACCTGTAA 229946

278 leargprogluGlu .....PheIleLysAspGluSerGlyAspIleGlu 292

229947 TACGCCCAAGAGATGTAAACCTCTTCCAAAGAAAGCAATTTTAAGC 229996

293 G1YthrIleArgAspSerValTYrLeuGlyLeuAsnThrAspTYrPhe 309

229997 GGAACCTTAACATCAACGCAATTTTCAAGAGCTTCATACCAATGACTCT 230046

309 egluThrGlyPheAlaSerLysIleGlnValSerGluGluSerThrPhe 326

230047 AGAA ..... 230050

326 luGluAspLeuGlnLysGlyAsnArgIle .....ArgLeuArg 338

230051 .....ATCCAAAAAACAATGTGATGTTCAAAAGCACAGACCTTACA 230092

339 IleAsnThrGlnLysLeuAsnIlePhe 347

230093 AAGCTTGAGACAGACGTTGATATTTT 230119

seq\_name: /std1/gcdata/geneSeq/geneSeqn-emb1/NA1996.DAT:AA142063

seq\_documentation\_block:

ID AA142063 standard; DNA; 1830121 BP.

XX AA142063;

XX 14-SEP-1999 (first entry)

XX Haemophilus influenzae complete genome sequence.

KW Genome; bacterium; Haemophilus influenzae; computer readable medium;

KW expression modulating fragment; regulation; gene expression; vector;

OS Haemophilus influenzae.

XX WO9633276-A1.

XX 24-OCT-1996.

XX 22-APR-1996; 96WO-US05320.

XX 07-JUN-1995; 95US-0487429.

XX 21-APR-1995; 95US-0426787.

XX 07-JUN-1995; 95US-0476102.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UYO) UNIV JOHNS HOPKINS.









318 nvalSerGlucInuSerThrPheGluGluAspLeuGlnLysGlyAsnArgI 335  
|||||  
940 .....TGGCAGCGGAGAGAGA 955  
335 leArgLeuArgIleAsnThrGlnLysLeu 344  
|||  
956 TATTATTGACGAGTCAACGCTACGCGTCTG 984  
seq\_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41224  
seq\_documentation\_block:  
ID AAH41224 standard; DNA; 349980 BP.  
AC AAH41224;  
XX  
XX 29-OCT-2001 (first entry)  
XX  
XX Pyrococcus abyssi genomic fragment #3.  
DE  
XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.  
XX  
XX Pyrococcus abyssi.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..49980  
FT /tag= a  
FT /note= "This sequence overlaps with the 3' end of  
FT AAH41223"  
FT misc\_feature 300001..349980  
FT /tag= b  
FT /note= "this sequence overlaps with the 5' end of  
FT AAH41225"  
XX  
XX FR2792651-A1.  
XX  
XX 27-OCT-2000.  
XX  
XX 21-APR-1999; 99FR-0005034.  
XX  
XX 21-APR-1999; 99FR-0005034.  
XX  
XX 21-APR-1999; 99FR-0005034.  
XX  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX (IFRE-) IFREMER INST FR RECH EXPL MER.  
XX  
XX Forlerre P, Thierri JC, Prieur D, Dietrich J, Lecompte O;  
XX Querellou J, Weissenbach J, Saurin W, Hellig R;  
XX  
XX WPI; 2001-126236/14.  
XX  
XX New nucleotide sequences isolated from Pyrococcus abyssi encode  
XX proteins useful in industry -  
XX  
XX Claim 1: Page 347-443; 1657pp; French.  
XX  
XX The present invention relates to the genomic sequence of Pyrococcus  
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a  
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal  
XX vents. The present sequence is a fragment of the genomic sequence of P.  
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223  
XX and the 3' end of this sequence overlaps with the 5' end of AAH41225. The  
XX proteins of the present invention have various potential industrial uses,  
XX since the proteins are stable at very high temperatures, some up to 110  
XX degrees centigrade.  
XX Note: This patent is in the same patent family as WO200065062, which  
XX contains additional sequences as shown in AAB99132-AAB99143,  
XX AAH75903-AAH75920 and AAG66436.  
XX  
XX Sequence, 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;  
SQ

alignment\_scores:  
Quality: 630.00 Length: 320  
Ratio: 2.751 Gaps: 5

Percent Similarity: 71.562 Percent Identity: 43.125  
alignment\_block:  
US-09-769-787-162 x AAH41224 ..  
Align seg 1/1 to: AAH41224 from: 1 to: 349980  
22 leuAsnIleThrIleProLysGlySerLeuPheThrLeuGluAlaIle 38  
|||||  
154056 GTTACCTTGAGCTTAAGGACGCGAGATTATTCATCCCTGGGCCGAG 154105  
38 rGlyCysGlyLysThrPheLeuArgMetIleAlaGlyPheAsnSerI 55  
|||||  
154106 CGGTTGGCGCAAGACGACGCTCAGGATATGCGCGTTGGAAAGCC 154155  
55 leuGluGlyGluPheThrPheAspThrLysIleAsnMetCysIu 71  
|||||  
154156 CAGACAAAGAAATGCTGTTATTCAGTGTAGGATGATGATCTTGTGCC 154205  
72 ProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIlePhePr 88  
|||||  
154206 CCTTACTCGAGAAACATAGCGCTGCTCCAGGACTACGCTTGTGCC 154255  
88 OhIleuThrValArgAspAsnValAlaPheGlyLeuMetClnLysLysV 105  
|||||  
154256 ACATATGAACTTTTCAGAAAGCTGGCTTGGCTTGAGGTTAGGAAGC 154305  
105 alProLysGluGluLeuIleGlnInhrAsnLysTyrLeuGluMet 121  
|||||  
154306 TTCCAGCAAGACATAGACGAGAGGTTAGGAACTTTGAAACTGTT 154355  
122 GlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyGlyI 138  
|||||  
154356 GCTTGAAGGCTTGTGAAATCGCATGCCGAAACATTAACCGAGGCTCA 154405  
138 nGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAspProSerValI 155  
|||||  
154406 ACAACAGAGGCTGCTTGCCTGAGCGCTTGTATGAAACCGAGGCTCC 154455  
155 eulMetAspGluProLeuSerAsnLeuGluAlaLysLeuArgLeuAsp 171  
|||||  
154456 TCCTCTTAGAGCAACCTTAAGCAATCTAGATGCCCAAGTTAGAGAGCC 154505  
172 MetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrIle 188  
|||||  
154506 CTGAGATGCGAGATTAAGAAATACAGAGAGGTTGGCATTAAGAGCAT 154555  
188 lTyrValThrHisAspGlnGluAlaMetAlaIleSerAspGlnIleA 205  
|||||  
154556 GTACGTTACCCATGACCAAGAGGCCATGCTATTAAGCAGCAATAG 154605  
205 laValMetLysAspGlyValIleGlnGlnIleGlyArgProLysGluLeu 221  
|||||  
154606 CGGTCATGAACTTTCGGAAGATTAAGCAGCTTGAAGGCCCTAGACTTC 154655  
222 TyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgThrAs 238  
|||||  
154656 TACTAACACCGAGAGGAGCTGCTCCAAATTCTCGGAACGGGAAA 154705  
238 nIleIleProAlaAsnLeuGluLysArgSer..... 248  
|||||  
154706 TTTCG.....CTAAAGCTCCCTTCACGACGCGCTTCATGCT 154743  
249 .....aspGlyAlaTyrIleValPhe 255  
|||||  
154744 TGGGAGAGCTCTGCTTCAAGCTCGGGGTGATGGGCTGTAAAGTGTTC 154793  
256 SerAspGlyTyrAlaLeuArgMetProAlaLeuAspGlnValGluGluI 272  
|||||  
154794 TTATGCGCTGAAGAGCTTGAAGATA...GGAAGAGAGGCGTTGAGCTGA 154840  
272 nAlaIle.....HisValSerI 278  
|||||

154841 GGTATGAGCTACGAAATACCTCCCGGAGATAGTTAGCTTAGCG 154890  
278 leaIrProGluGluPheIleLysAspGluSer..GlyAspIleGluGlyT 294  
154891 TTCAAGGTAAAGAAATATATAGCGGAGACTCTTGGCAAGATGAGAG 154940  
294 hrIleArGAspSerValTyrLeuGlyLeuAsnThrAspTyrPheIleGlu 310  
154941 ATACGAGAGAAAGTTATATATAAAGTGAAGAGCTCAGCTTCTATAGAC 154990  
311 ThrGlyPhe 313  
154991 TCTGGATTTC 154999  
seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546248  
seq\_documentation\_block:  
ID AA546248 standard; DNA: 14674 BP.  
XX  
AC AA546248;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE DNA encoding novel mar regulated protein (NIMR) #17.  
XX  
KM mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.  
XX  
OS Escherichia coli.  
XX  
PN W0200170776-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 08-MAR-2001; 2001WO-US07478.  
XX  
PR 10-MAR-2000; 2000US-188362P.  
XX  
PA (TUFT ) TUFTS COLLEGE.  
XX  
PI Levy SB, Barbosa TM, Aleksnun MN;  
XX  
DR WPI: 2001-602769/68.  
DR P-PSDB: AAU29348.  
XX  
PT Identifying compounds that modulate a newly identified mar regulated  
PT polypeptide activity, useful as antimicrobial compounds, involves  
PT contacting the polypeptide with a test compound .  
XX  
PS Disclosure; Page 266-276; 526pp; English.  
XX  
CC The invention relates to a method of identifying compounds that modulate  
CC a newly identified mar regulated (NIMR) polypeptide activity. The method  
CC comprises contacting an NIMR polypeptide with a test compound under  
CC interaction conditions, determining the ability of the compound to  
CC modulate the activity or expression of the polypeptide, and selecting the  
CC modulators. NIMR nucleic acids and polypeptides are used in the treatment  
CC of microbial infections, and in screening for modulators of NIMR  
CC expression and activity. These modulators can be used to reduce the  
CC infectivity of a microbe on a surface, and the virulence of a microbe in  
CC a subject suffering from an infection. AA546232-AA546278 represent  
CC Escherichia coli NIMR coding sequences of the invention.  
XX  
SQ Sequence 14674 BP; 3637 A; 3767 C; 3842 G; 3428 T; 0 other;

alignment\_scores:  
Quality: 628.00 Length: 358  
Ratio: 2.482 Gaps: 6  
Percent Similarity: 70.670 Percent Identity: 36.872

alignment\_block:  
US-09-769-787-162 x AA546248 ..

Align seg 1/1 to: AA546248 from: 1 to: 14674  
1 MetSerGluIleLysIleIleAsnAlaLysLysIleTyrHisAspValPr 17  
13495 ATGGGAGAGGTACAGCTGCMAAATGTAAACGAAGCCGTGGGAGAGTCT 13544  
17 oValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuPheThrL 34  
13545 GGATTCGAAAGATATCATATCTGATATCCATGAAGGTGAATTCGTGTCT 13594  
34 euleuGlyAlaSerGlyCysGlyLysThrThrLeuLeuArgMetIleAla 50  
13595 TTGTGAGACGCTGTGCTGCGTAATGACATCTTACTGGCGATGATGGCC 13644  
51 GlyPheAsnSerIleGluGlyGlyLysPheTyrPheAspThrLysIle 67  
13645 GGGCTTGAGACGATCACCGCGCGCCGCTTCATCGGTGAGAAACGGAT 13694  
67 eAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAsnT 84  
13695 GAATGACACTCCGCCAGCAGAACCGCGCTGTGTATGTGTTCAGTCTT 13744  
84 YrAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyLeu 100  
13745 ACGCGCTATATCCACCTGATAGCAGAAACATGTCATTTGGCGCTG 13794  
101 MetGlnLysLysValProLysGluGluLeuIleGlnGlnThrAsnLysTy 117  
13795 AAACGTGGCTGGCGCAAAAAGAGTGATTAACCAACGGGTAAACCGGT 13844  
117 rLeuGluLeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysL 134  
13845 GCGCGAAGTGTCTACAACTGCCCATTTGCTGATGCTCAAAACCGAAAGCC 13894  
134 euSerGlyGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaVal 150  
13895 TCTCGGTGTGTCAGCGTCAGCGTGGCGATTTGGCGCTGAGTGGGCC 13944  
151 AsnProSerValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaL 167  
13945 GAGCCAAAGGTATTTTGTCTGATGAACCGCTCTCCAACTGATGCTCTC 13994  
167 sLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluVal 184  
13995 ACTGCGTGTGCAAAATGCTGATGAAATCTCCGCTGCTAAACGCTGTG 14044  
184 lYleThrThrValTyrValThrHisAspGlnGluAlaMetAlaIle 200  
14045 GCCGCACANTGATTTTACGTACCACAGATGAGTGCAGATGAGCGCTG 14094  
201 SerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyAr 217  
14095 GCCGCAAAATGCTGTGCTGAGACCGCGGTGCGCTGCCAGAGTTGGAA 14144  
217 gProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheI 234  
14145 ACCGCTGGAGCTGTACCACTATCCGACACCGCTTTTGTGCGCGATTAA 14194  
234 lGlyLys.....ArgThrAsnIleIleProAlaAsnLeuGluLysArgSer 248  
14195 TCGGTTCCGCCAANAAGATGAACTTCTGCGGTAA..... 14229  
249 AspGlyAlaTyrIleValPheSerAspGlyTyrAlaLeuArgMetProAl 265  
14230 .....GTGACCGCCACCGC 14243  
265 AlLeuAspGlnValGluGluGlnAlaIleHisValSerIleArgProGluG 282  
14244 AATCGATCAAGTGCAGGTGAG...CTGCCGATGCCAAATGCTGACAGAG 14290  
282 lPheIleLysAspGluSerGlyAspIleGluGlyThrIleArgAspSer 298  
14291 TCTGGCTGCCAGTTGAAAGCCGATGATGTCAG.....GTGAGAGCAAT 14334



```

280 roglu...PheIleLysAspGluSerGlyAspIleGluGly...Thr 294
      ||||| ..... ||||| ..... ||| :
,839 CAGAACATTTTATCTTGCTGCGAGAGGATGCGATGCACAACTTGGCAG 888
295 IleArgAspSerValTyrLeuGlyLeuAsnThrAspTyrPheIleGluThr 311
      ||||| ..... ||||| ..... ||| :
889 ATTAAGAAGCGGTTATATGGGCAACCATTTGGGAGAGTGTTCGAAC 938
311 rGIpheaIleSerLysIleGluValSer 320
      ||||| ..... ||| :
939 TGGCGGGCGCAAGATTACTGTAAAT 966

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA546257
seq_documentation_block:
ID   AA546257 standard; DNA; 11756 BP.
XX
AC   AA546257;
XX
DT   18-DEC-2001 (first entry)
XX
DE   DNA encoding novel mar regulated protein (NIMR) #26.
XX
KN   mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200170776-A2.
XX
PD   27-SEP-2001.
XX
PE   08-MAR-2001; 2001WO-US07478.
XX
PR   10-MAR-2000; 2000US-188362P.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Levy SB, Barbosa TM, Aleksun MN;
XX
DR   WPI: 2001-602769/68.
XX
DR   P-PSDB; AAU29357.
XX
PT   Identifying compounds that modulate a newly identified mar regulated
PT   polypeptide activity, useful as antimicrobial compounds, involves
PT   contacting the polypeptide with a test compound .
XX
PS   Disclosure; Page 344-351; 526pp; English.
XX
CC   The invention relates to a method of identifying compounds that modulate
CC   a newly identified mar regulated (NIMR) polypeptide activity. The method
CC   comprises contacting an NIMR polypeptide with a test compound under
CC   interaction conditions, determining the ability of the compound to
CC   modulate the activity or expression of the polypeptide, and selecting the
CC   modulators, NMR nucleic acids and polypeptides are used in the treatment
CC   of microbial infections, and in screening for modulators of NIMR
CC   expression and activity. These modulators can be used to reduce the
CC   infectivity of a microbe on a surface, and the virulence of a microbe in
CC   a subject suffering from an infection. AA546232-AA546278 represent
CC   Escherichia coli NIMR coding sequences of the invention.
XX
SQ   Sequence 11756 BP; 2739 A; 2892 C; 3160 G; 2965 T; 0 other;

alignment_scores:
      Quality: 627.00      Length: 346
      Ratio: 2.488      Gaps: 5
Percent Similarity: 72.832      Percent Identity: 39.017

alignment_block:
US-09-769-787-162 x AA546257      .
Align seg 1/1 to: AA546257 from: 1 to: 11756

```

```

4   IleLysIleIleAsnAlaLysLysIleTyrHisAspValProValIleGI 20
      ..... ||| ||| ||| .....
7032 TTAGAAATCCGCAACTGCACCAATCTTACGATGCTCAACATCGGTGG 7081
20   uAsnLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGlyA 37
      ..... ||| ||| ||| .....
7082 TGATGTCAGCGTCCATCATTCACAAAGTGAAATCTTCGGCGTCTGGGG 7131
37   lAserGlyGlyLysThrThrLeuLysArgMetIleAlaGlyPheAsn 53
      ||||| ..... ||||| ..... ||| :
7132 CATCCGCGTGTGGCAATCCACCTCTGCTATGCTGGCAGATTTCGAA 7181
54   SerIleGluGlyGlyLupheYrPheAspPheThrLysIleAsnSme 70
      ..... ||| ..... ||| .....
7182 CAACCTTCTGCGGACAGATATGCTTGAATGGCTGATTTGTCACAGGT 7231
70   tGIuProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIleP 87
      : ||| ||| ..... ||| ..... ||| :
7232 TCCGCTTACCTGCGCCCATCAATATGATGTTTCACTTACGCGCTGT 7281
87   heProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
      ||||| ..... ||||| ..... ||| :
7282 TTCCGATATGACCGTGGAAACAGACATCGCTTGGCTGGAACACAGAC 7331
104   LysValProLysGluGluLeuIleGlnInThrAsnLysTyrLeuGlu 120
      ||||| ..... ||||| ..... ||| :
7332 AAACCTACCGAAAGCGAATGCGCAGCGGTCAATGAGATGCTCGGCT 7381
120   uMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyG 137
      ||||| ..... ||||| ..... ||| :
7382 GGTGATATGACAGAGTTGCGCAACGCAACCGCATTCGCGGTG 7431
137   LysGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSer 153
      ||||| ..... ||||| ..... ||| :
7432 GTCAGCGACAACTGTGGCTGCGCCCGCAAGCCTTCGGAAGCCCGGAA 7481
154   ValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArg 170
      ..... ||||| ..... ||||| .....
7482 CTATTTACTGCTGATGACCGATGGCGCGCTGATTAATAAAGCTGGTGA 7531
170   uAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThr 187
      ..... ||| ..... ||| .....
7532 CAGATGACAGCTTGAAGTGTGATTTCTGAGCGCGTGGTGACTT 7581
187   hrValTyrValThrHisAspGlnGluAlaMetAlaIleSerAspGln 203
      ||| ||||| ..... ||||| ..... ||| :
7582 GTGTGATGTGCACCCACGATCAGGAAGAGCGATGACATGCGGGGCGC 7631
204   IleAlaValMetLysAspGlyValIleGlnInIleGlyArgProLysGI 220
      ||||| ..... ||| ..... ||| .....
7632 ATGCGCATTTATGAATGCTGGGAAATTTGTCACAAATGGCAACCGGAAG 7681
220   uLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
      ||||| ..... ||| ..... ||| .....
7682 GATCTACGACATCCGACTACCGCTATAGCCGGAATTTATATGGCTCG 7731
237   hrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrIle 253
      ||||| ..... ||| ..... ||| .....
7732 TAAACGCTTTGAAGCGTACTCAAGAGCGTCAGGAAGATGCGCTGTG 7781
254   ValPheSerAspGly.....TyrAlaLeuArgMetProAla...LeuAs 267
      : ||| ||| ..... ||| ..... ||| .....
7782 CTGTATTCGCGCGGCGTGTGATCCATCAAGATGCGACCGGATGCTTC 7831
267   pGIuValGluGluGlnAlaIleHisValSerIleArgProGluGluPheI 284
      : ||||| ..... ||||| ..... ||| :
7832 GGTGTCGATTAACGTCCGCTACATGTGCGCGCCCGGAAATAATCA 7881
284   lLysAspGluSerGlyAspIleGlu.....GlyThr 294
      : ||||| ..... ||| .....
7882 TGCTTTGCGAAAGCCGCCCGCAATGATTGTAATTCGCGGTGGGGAG 7931

```











```

470 CGTCCGGGTGGGAAGACGACACCGATGGCGATGATGACGGCGAFTGCAG 519
54 SerIleGIuGIyglYgluPheTyTrPheAspSerThrIysIleAsnAsmE 70
   ::::  |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
520 ACCCCGACGTAAAGGGGATCCGGCTCGAAGGACCCGACGTGTGCAGAC 569
   ::::  |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
70 tGIuProSerLysArgAsnIleGIyMeTValPheGIAsnIlyrAlaIleP 87
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
570 CCCACCCAAACAGCGCACTCAACACGGTGTTCACGACTACCGCGCT 619
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
87 hebProHisLeuThrValArgAspAsnValAlaPheGIyLeuMeTcIuLys 103
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
620 TCCCGACATAGAGGCTGTGGGACCAACTCGGTGACGGCCCGACAGCAAG 669
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
104 LysValProLysGIuGIyLeuIleGIeGIuThrAsnIlyrGIyGIuLe 120
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
670 AAATCTGGCAAAAGCGAGTCCGCAACCGCGGTCAACGACGTGTGGAGAT 719
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
120 uMeTGIuIleAlGIeGIyrAlaAspArgLysProAspLysLeuSerGIy 137
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
720 CGTCCGGGTGACCGGAATTTGCCAGGCGCAGGCCGCCCGCCAGCTGTCCGG 769
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
137 LysIleGIuGIuArgValThrLeuAlaCysAlaLeuAlaValAsnProSer 153
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
770 GGCAGACAGCAGCGGTGGCTGGCCCGGACACTGTGAATACCCACAGC 819
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
154 ValLeuIleuMeTAspGIuProLeuSerAsnIleuGIuAlaLysLeuArg 170
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
820 GCGCTGTGTCTCAATGAACCGCTGGAGCTGCACCTCAACGAAAGCTGGCCA 869
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
170 uAspMetArgGIuAlaIleArgGIuIleGIuIleGIuValGIuLysIleThr 187
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
870 CGTCAATGCAATTCGAGCTCAAGCCATCCACGGGAGGTCCGGATCAAGT 919
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
187 hValIlyrValIThrHisAspGIuGIuAlaMeTAlaIleSerAspGIu 203
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
920 TCACTTACGTGACCCACGACGAGAAAGGCGCTCACATGATGACGCCGC 969
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
204 IleAlaValMeTAspArgIyValIleGIeGIuIleGIyArgProLysGI 220
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
970 ATGCGGTGTGAACCGCGGACGCTGAACATGATGGCAGCCGACCCGACCGA 1019
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
220 uLeuTyTrHisLysProAlaAsnGIuPheValAlaThrPheIleGIyArgT 237
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1020 GATCTACGACCGTCCGCGAGAGGTTGTCGTCCAGGCTTATCGGACAGG 1069
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
237 hAsnIleIleProAlaAsnLeuGIuLysArgSerAspGIyAlaTyrlle 253
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1070 CCAACCTCTGGGGGCGGTGCACCGCGCGCTCAACCGGATTAATCTC 1119
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
254 ValPheSerAsp...GIyTrAlaIleArgMeTProAlaLeuAspGIuVal 269
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1120 GAATCGACACTTCTCGGCTCGAGCTTAAGGCACGCCCGGCGAGACCAAC 1169
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
269 IGIuGIuGIuAlaIleHisValSer.....IleArgProGIuGIuPhe 283
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1170 GATCGAGCCCGCGGCGGACCGACCTGATGTGCTGCCGGAACGACATCC 1219
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
284 .....IleLysAsp...GIuSerGIyAsp.....Ile 291
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1220 GGGTCAACCCCGGCTCCAGACGCGGACCGGACCGGTGACGTGCGCTGCTG 1269
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
292 GIuGIyThrIleArgAspSerValTyTrLeuGIy 302
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
1270 CGTGCACCGCTCAACGACCTGACCTTCCAAAGT 1302
   |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::
seq_name: /cgn2.6/ptodata/1/lna/5B_COMB.seq:US-08-997-362-88

```

APPLICANT: Hiyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970  
FILING DATE: June 12, 1997  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
FILING DATE: August 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 88:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1518 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-997-362-88

```
seq_documentation_block:
; Sequence 88, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
```

```

alignment_scores:
    Quality: 610.00
    Ratio: 2.652
    Percent Similarity: 73.955
    Percent Identity: 42.444

alignment_block:
US-09-769-787-162 x US-08-997-362-88
..

Align seg 1/1 to: US-08-997-362-88 from: 1 to: 1518

4 IIElyIlelleasnaIalysIleIyrHIsaPValProvalIlegl 20
|||||:|||| :||||: ||| :||| :||||:
370 ATCGAGATCGACCAATCGATCGACGAAGCGCTTGGCGCAGATCACTGCGCGC 419
.. :|||:||||| |||:||||:|||||
420 GGACGACGACTTCTCATCGCGCCCGGGGAGATTCTTCCATCGCTCGGCC 468
.. :|||:||||| |||:||||:|||||
420 uasleuIasnIlethrIleProIySglySerIeupheThrIeuleuIyla 37
.. :|||:||||| |||:||||:|||||
37 IasSerGIyCSglIySthrIhrIeuleuAargMetIleIagIyPhasn 53
|||:|||||:||||| |||:||||:|||||
470 CCGCGGGGTGGGAAGACGACCAAGCTGGCCAGATCGCGGGGATTCGAG 519
54 SerIleIagIyGIyIupheIyrPheaspsaphrIyIleasnaIame 70
||| :||| :||| :|||:||||:
520 ACCCCGACTGAAGGGGCGCATTCGCCCTCGAAGCCGCCGACGCTGTGAAGAC 565

```

```
70 tgluProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIleP 87
|||||
570 CCACCCACCAAGCGCAACGCTCAACAGCGGTCTCCAGCACTACGCGCTCT 619
87 heProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
|||||
620 TCCCGCATGACGCTGTGGGACACGCTGCTACGCGCGCGCGGACGCAAG 669
104 LysValProLysGluGluLeuIleGlnGlnInThrAsnLysTyrLeuGlu 120
|||||
670 AAACCTCGCAAGCGGAGGTCCGCAAGCGCTCGAGAGCTGCTGGAGAT 719
120 uMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyG 137
|||||
720 CCTCGCGGTGACGGAATTTGCCGAGGCGCGCCCGCCAGCTGTCCGGGG 769
137 LysGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSer 153
|||||
770 GGCACACGACGCGGTGGCGGTGGCGCGGCGCTGCTGAACCTACCCAC 819
154 ValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArgLe 170
|||||
820 GCGCTCTCTCTGATGACCGCTCGAGCGCTCGACCTGACGCTGCGCCA 869
170 uAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrT 187
|||||
870 CCGTACATGAGTTCGAGCTCAGCGGATCCAGCGGGAGGTGCGGATCAG 919
187 hValIleTyrValThrHisAspGlnGluAlaMetAlaIleSerAspGln 203
|||||
920 TCATCTACGTCAGCCACGACGAGGAGGCGCTCAGATGAGTGACCCG 969
204 IleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
|||||
970 ATCCGCGGTATGACCGCGGCAACGCTCGACAGATCGGACCGCCGCA 1019
220 uLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
|||||
1020 GATCTACGACCGTCCCGGACGCGTCTCGCGACCTTCATCGGACGAG 1069
237 hAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrIle 253
|||||
1070 CCACCTCTGGGCGGCGCTGACCGCGGCTCCCAACCGCATTTACGTC 1119
254 ValPheSerAsp...GlyTyrAlaLeuArgMetProAlaLeuAspIle 269
|||||
1120 GAGATCGACGTTCTCGGCTCGAGCTGAAAGGACACCGCGGCGAGCCAC 1169
269 IgluGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
|||||
1170 GATCGAGCGCGGCGGCGGCGGACCGCTGATGGTGGCTCGGGAACGATCG 1219
284 .....IleLysAsp...GluSerGlyAsp.....Ile 291
|||||
1220 GGGTCAACCCGGGCGCTCCAGAGCGCGCGACGCTGACGCTCGCGCTG 1269
292 GlnGluThrIleArgAspSerValTyrLeuGly 302
|||||
1270 CGTCCGACCGCTACCGACCTGACCTTCCAAGGT 1302
seq. name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:us-08-873-970-88
seq. documentation block:
: Sequence 88, Application US/08873970
: Patent No. 6001361
: GENERAL INFORMATION:
: APPLICANT: Tan, Paul
: APPLICANT: Hiyma, Jun
: APPLICANT: Visser, Elizabeth
: APPLICANT: Skinner, Margot
: APPLICANT: Scott, Linda
: APPLICANT: Prestidge, Ross
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
```

```

: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
: NUMBER OF SEQUENCES: 106
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Law Offices of Ann W. Speckman
: STREET: 2601 Elliott Avenue, Suite 4185
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/873,970
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/705,347
: FILING DATE: 29-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Sleath, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000.1002C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
: TELEX:
: INFORMATION FOR SEQ ID NO: 88:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1518 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-873-970-88

alignment_scores:
: Quality: 610.00 Length: 311
: Ratio: 2.652 Gaps: 5
: Percent Similarity: 73.955 Percent Identity: 42.444

alignment block:
US-09-769-787-162 x US-08-873-970-88 ..

Align seg 1/1 to: US-08-873-970-88 from: 1 to: 1518

4 IleLysIleIleAsnAlaLysLysIleTyrHisAspValProValIleG 20
|||||
370 ATCGAGATGACCATGTCTCAGAAAGCGCTCGGCGACTTACGTGCGCTCG 419
20 uAsnLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGluYA 37
|||||
420 GGCAGGAGACTTCTCCATCGCGCGCGGAGGTCTTCTCCATGCTCGGCG 469
37 IAserGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsn 53
|||||
470 CGTCCGCGGTGGGAGACGACGACGCTGCGCATGATCGCGGATTCGAG 519
54 SerIleGluGlyLysLysPheAspPheAspThrLysIleAsnAsnHe 70
|||||
520 ACCCGCATGAAAGGCGCATCCGCTCGAAAGCGCGGACGTGCGAGGAC 569
70 tgluProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIleP 87
|||||
570 CCACCCACCAAGCGCAACGCTCAACAGCGGTCTCCAGCACTACGCGCTCT 619
87 heProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
|||||
620 TCCCGCATGACGCTGTGGGACACGCTGCTGACGCGCGCGGCGGACGCAAG 669
```

```

104 LysValProLysGluLeuIleGlnGlnThrAsnLysTyrLeuGlu-120
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
670 AAACCTGGCAAGACGCGATCCCAACGCGCTGCACGACGCTGCGAGAT 719
120 uMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGly 137
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
720 CGTCCGGCTACCGAATTTCGCCGACGCGACCGCCCGCCACTGTCGCGCG 769
137 LysGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSer 153
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
770 GCGACGACGACGCGGTGCGTGGCCCGGCGACGCTGCACTACCCAGC 819
154 ValLeuLeuMetAspGluProLeuSerAsnLeuGlnAlaLysLeuArg 170
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
820 GCGCTGCTGCTGATGAACCGCTCGGAGCGCTGACCTGAAGCTGCGCCA 869
170 uAspMetArgGlnAlaIleArgGluIleGlnIleGlnValGlyLeuThr 187
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
870 CGTCATGCACTTGCAGCTCAAGCGCATCCAGCGGAGGTGCGATCACGT 919
187 hValTyrValThrHisAspGlnGluGlnAlaMetAlaIleSerAspGln 203
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
920 TCATCTACGACGACCAACGACGACGAGAGCGCTCAGATGAGTACGCG 969
204 IleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
970 ATCGCGGTGATGAACCGCGCAACGCTGCAACAGATCGGACCGCGAGA 1019
220 uLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArg 237
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1020 GATCTACGACGCTGCGCGACGCGGTGCTGTCGCGCATCTGATCGGAC 1069
237 hAsnIleIleProAlaAsnLeuGlnLysArgSerAspGlyAlaTyrIle 253
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1070 CCAACCTCTGGCGGCGCGGTGCACGCGCGCTCCCAACCGCGATTAAGT 1119
254 ValPheSerAsp...GlyTyrAlaLeuArgMetProAlaLeuAspGlnVa 269
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1120 GAGATGACGCTTCTCGCTCGACGCTGAAGGACGCGCGGCGGAGACAC 1169
269 IGIuGlnGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1170 GATCGACCGCGCGGCGCGACCGCATGATGTCGCTCCGGAACGCACTCC 1219
284 .....IleLysAsp...GluSerGlyAsp.....Ile 291
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1220 GGGTCACCGCGGCGCTCCCGACGACGCGCGCGCTGACGCTGCGCTGCG 1269
292 GluGlyThrIleArgAspSerValTyrLeuGly 302
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1270 CGTCCGACCGCTCACCGCATGACCTTCCAAAGT 1302
seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:us-09-095-855-88
seq_documentation_block:
; Sequence 88, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Compounds and Methods for
; TREATMENT OF INFECTIONS: Treatment and Diagnosis of Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; *ADDRESS: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1518 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-095-855-88

alignment_scores:
      Quality: 610.00      Length: 311
      Ratio: 2.652
      Percent Similarity: 73.955      Percent Identity: 42.444

alignment_block:
US-09-769-787-162 x US-09-095-855-88 ..
Align seg 1/1 to: US-09-095-855-88 from: 1 to: 1518

4 IleLysIleIleAsnAlaLysLysIleTyrHisAspValProValIleG 20
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
370 ATCGAGATCCACCATGTCACGAAGCGCTGCGGACATCTGCTCCATGCTCG 419
20 uAsnLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGluVa 37
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
420 GGAAGCAGACTTCTCCATGCGCGCGCGGAGGATCTTCTCCATGCTCGCC 469
37 IAspGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsn 53
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
470 CGTCCGCGTGTGGAGACGACGACGCTGCGCATGATGCGGGGATTCGAG 519
54 SerIleGluGlyGluPheTyrPheAspAspThrLysIleAsnSme 70
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
520 ACCCGCACTGAAGGGGATCCCGCTCGAAGCGCGCGCTGTCGAGAC 569
70 tGluProSerLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIle 87
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
570 CCCACCCAAACAGCGCAAGTCACACGAGGTTCACGACTACGCTGCTGT 619
87 hProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnLys 103
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
620 TCCCGCAGATGACGCTGTGGACACAGCTGCGCGTACGCGCCGCGACAG 669
104 LysValProLysGluLeuIleGlnGlnThrAsnLysTyrLeuGlu 120
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
670 AAACCTGGCAAGACGCGATCCCAACGCGCTGCACGACGCTGCGAGAT 719
120 uMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGly 137
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```
720 CGTGGCGCTGACCGAATTTGCCAGCGAGCCGCCCGCAGCTGCGCGCG 769
437 IYGLGNGInIaRgYalThrLeuAlaCyAlaLeuAlaValaInProSer 153
770 GGCAGAGCAGCGGGGTGGGCGCGCGCGCTGCTGAACACTACCCAGC 819
154 ValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaValaLysLeuArg 170
820 GCGGCTGCTCTCGATGAACCGCTCGAGCGCTCGACCTGAACCTGCGCCA 869
170 uAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrT 187
870 CGTCAATGCGTTCGAGCTCAACGCCATCCAGCGGAGGTGGGATCACGT 919
187 hValTyrrValThrHisAspGlnGluAlaMetAlaIleSerAspGln 203
920 TCATCTACGTGACCCAGCAGACAGAGAGCGCTCACGATGAGTGCAGCGC 969
204 IleAlaValaMetLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
970 ATCGGCGGTATGACGCGCGCAACGTGCAACAGATCGCAGCGCGACCGCA 1019
220 uLeuTyrrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
1020 GATCTACGACCGCTCCCGCAGCGGTGTTCTCGCCAGCTTCATCGCAGAG 1069
237 hAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrrIle 253
1070 CCAACCTCTGGGCGGCGGCTGCACCGCGCGCTCCAAACCGCGATTCAGTC 1119
254 ValPheSerAsp...GlyTyrrAlaLeuArgMetProAlaLeuAspGlnVa 269
1120 GAGATCGACGCTCTCGCTCGACGCTGAAAGCAGCGCGCGGCGAGACAC 1169
269 IeGluGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
1170 GATCGAGCGCGCGGCGGCGCACCGCTGATGGCGTCCGGAACGCGATCC 1219
284 .....IleLysAsp...GluSerGlyAsp.....Ile 291
1220 GGGTCAACCGCGGCTCCAGAGCGCGCGCGAGTGCATCGCTCGCGTGG 1269
292 GluGlyThrIleArgAspSerValTyrrLeuGly 302
1270 CGTGCACCGCTCACCGACCTGACCTTCCAAGGT 1302
seq_name: /cgn2_6/ptodata/1/ina/68_COMB.seq:US-09-324-542-88
seq_documentation_block:
: Sequence 88, Application US/09324542
: Patent No. 6328978
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Tan, Paul L.J.
: APPLICANT: Prestidge, Ross
: TITLE OF INVENTION: Methods and Compounds for the Treatment
: TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
: FILE REFERENCE: 11000.1007c1
: CURRENT APPLICATION NUMBER: US/09/324,542
: EARLIER FILING DATE: 1999-06-02
: EARLIER APPLICATION NUMBER: US 08/997,080
: EARLIER FILING DATE: 1997-12-23
: NUMBER OF SEQ ID NOS: 194
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 88
: LENGTH: 1518
: TYPE: DNA
: ORGANISM: Mycobacterium vaccae
US-09-324-542-88
alignment_scores:
Quality: 610.00 Length: 311
```

```
Ratio: 2.652 Gaps: 5
Percent Similarity: 73.955 Percent Identity: 42.444
alignment_block:
US-09-769-787-162 x US-09-324-542-88 ..
Align seg 1/1 to: US-09-324-542-88 from: 1 to: 1518
4 IleLysIleIleAsnAlaLysLysIleTyrrHisAspValProValIleG 20
370 ATCGAGATGACCATGTCACGAAAGCGCTTCGGCGAGCTACCTGCGCGC 419
20 uAsnLeuAsnIleThrIleProLysGlySerLeuPheThrLeuGlyA 37
420 GGAAGCAGACTTCTCATGCGCGCGGAGAGTTCCTTCATGCTCGGCC 469
37 IeSerGlyCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsn 53
470 CGTCCGGGTGTGGAAAGACGACGATGCGCATGATCCGCGGATTCGAG 519
54 SerIleGluGlyGlyLupheryrPheAspAspThrLysIleAsnAsnMe 70
520 ACCCGGACTGAAAGGGGCGATCCGCTCGAAAGCGCGGAGCTGCGAGAC 569
70 tGluProSerLysArgAsnIleGlyMetValPheGlnAsnTyrrAlaIleP 87
570 CCCACCCAAAGCGCAACGCTCACACGSGTGTTCGACGACTACGCGCTGT 619
87 hProHisLeuThrValaArgAspAsnValaIlePheGlyLeuMetGlnLys 103
620 TCCCGCATGACGCGTGTGGACAAACGTCGCGNACGCGCGCGGCGACAG 669
104 LysValProLysGluLupheryrIleGlnGlnIleThrAsnLysTyrrLeu 120
670 AAACCTGCGAAAGCGCGATCCGCAAGCGCGTGCAGACCTGCTCGAGAT 719
120 uMetGlnIleAlaGlnTyrrAlaAspArgLysProAspLysLeuSerGlyG 137
720 CGTCCGGCTGACCGAATTTGCCGAGCGAGCGCGCGCGCGAGCTGTCCGGG 769
137 IYGLGNGInIaRgYalThrLeuAlaCyAlaLeuAlaValaInProSer 153
770 GGCAGAGCAGCGGGGTGGGCTGGCGCGCGGCGACTGCTGAACACTACCCAGC 819
154 ValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaValaLysLeuArg 170
820 GCGCTGCTGCTCGATGAACCGCTCGGAGCGCTGACCTGAACCTGCGCCA 869
170 uAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrT 187
870 CGTCAATGCGTTCGAGCTCAAGCGCATCCAGCGGAGTCCGGATCACGT 919
187 hValTyrrValThrHisAspGlnGluAlaMetAlaIleSerAspGln 203
920 TCATCTACGTGACCCAGCAGACAGAGAGCGGCTCACGATGAGTGCAGCGC 969
204 IleAlaValaMetLysAspGlyValIleGlnGlnIleGlyArgProLysG 220
970 ATCGCGGTATGAACCGCGCAACGTGCAACAGATGCGAGCGCGCGCA 1019
220 uLeuTyrrHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgT 237
1020 GATCTACGACGCTCCCGCAGCGGTGTTCTGCTGCGCAGCTTCATGCGACAG 1069
237 hAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrrIle 253
1070 CCAACCTCTGGGCGGCGGTGCACCGCGCGTCCAAACCGCGATTCAGTC 1119
254 ValPheSerAsp...GlyTyrrAlaLeuArgMetProAlaLeuAspGlnVa 269
1120 GAGATGACGCTTCTCGCTCGACGTGAAGCGACCGCGCGGCGAGACAC 1169
269 IeGluGlnAlaIleHisValSer.....IleArgProGluGluPhe. 283
```

```

1170 GATCGACCGCGCGGCGACCGCTGATGCTGCTCCGGAACGATCC 1219
284 .....llelyasp...Gluserglyasp.....ile 291
1220 GGGTACCCCGGCTCCAGACGCGCGGACGCTGACGCTGCTGCTG 1269
292 GluGlyThrIleArgAspSerValTyrIleuGly 302
      :::::|||||:::
1270 CGTGCCACCGCTACCGACCTGACCTTCCAAAGT 1302

```

seq\_name: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:US-09-381-862-5

```

seq_documentation_block:
; Sequence 5, Application US/09381862
; Patent No. 6245906
; GENERAL INFORMATION:
; APPLICANT: Ueyama, Hiroshi
; APPLICANT: Abe, Kanako
; APPLICANT: Keshi, Hiroyuki
; APPLICANT: Matsuhisa, Akio
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/381,862
; FILING DATE:
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 1997-71077
; FILING DATE: 25-MAR-1997
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP98/01288
; FILING DATE: 23-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Cawley, Jr., Thomas A.
; REGISTRATION NUMBER: 40,944
; REFERENCE/DOCKET NUMBER: 19036/36274
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; STRAIN: Clinical Isolate SP-26-46
; US-09-381-862-5

```

# alignment\_scores:

```

Quality: 601.00      Length: 301
Ratio: 2.671         Gaps: 3
Percent Similarity: 74.751      Percent Identity: 41.528

```

## alignment\_block:

US-09-769-787-162 x US-09-381-862-5 ..

Align seg 1/1 to: US-09-381-862-5 from: 1 to: 6688

```

8 AsnAlaLysLysIleTyrHisASP.....ValProValIleGluAsnLe 22
  ::::: ||| ::::: |||
3697 AATCTTAAAAAACATTTGAGACAGTGAACAGAGTCTTAAAGAACAT 3746
22 uasnIleThrIleProLysGlySerLeuPheThrLeuGluGlyAlaSerG 39
  ::::: ::::: ::::: ::::: ::::: :::::
3747 TAACTTGACCTTGAGAGAGAAATTTTACACCTGCTTGAGACCTTCTG 3796
39 LysCysGlyLysThrThrLeuLeuArgMetIleAlaGlyPheAsnSerIle 55
  ||| ::::: ::::: ::::: ::::: :::::
3797 GCCTTGGAATAATCACTATTAAATATATGCGGGGCTACTAGATGCC 3846
56 GluGlyGlyGluPheTyrPheAspAspThrLysIleAsnMetGluPro 72
  ::::: ::::: ::::: ::::: :::::
3847 AGCAGCGGACACATTTTATGATGGGGAACGTTATTAATGATTTGCCG 3896
72 OserLysArgAsnIleGlyMetValPheGlnAsnTyrAlaIlePheProH 89
  ::::: ::::: ::::: ::::: :::::
3897 TAATTAACGTGACATTCATCCGTTTCCAAATATATGCTTCCCTC 3946
89 IseuThrValAlaArgAspAsnValAlaPheGlyLeuMetGlnLysVal 105
  ||| ::::: ::::: ::::: :::::
3947 ATATGACTGTTTGGAGAAATGTCCTTGTGAGCTGAAGAAAGTG 3996
106 ProLysGluGluLeuIleGlnGlnThrAsnLysTyrLeuGluLeuMetG 122
  ||| ::::: ::::: ::::: :::::
3997 GACAGAAAGAGATGCTTAAGCGCGTGAAGAAACCTTGAAATGCTTCA 4046
122 nIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyGlyGlnG 139
  ::::: ||| ::::: ::::: :::::
4047 ATTAAGAGCGTATGAATAATGCTTCTATCCAGAAATATGCTGCGTCAAC 4096
139 GlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsnProSerValLeu 155
  ::::: ::::: ::::: :::::
4097 GTCAACGTGTTGCCATTCGCGCGTGCATATCAATCAACCGCTGTGTC 4146
156 LeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArgLeuAspMe 172
  ||| ::::: ::::: ::::: :::::
4147 TTACTTGACGACACACCTCAGCCCTTGATTTGAATTAAGACAGAGAT 4196
172 LArgGlnAlaIleArgGluIleGlnHisGluValGlyIleThrValT 189
  ::::: ::::: ::::: :::::
4197 GCAATATGATTAATTCGTAATTCGCAACACCTTACGATTAACCTTGTTT 4246
189 TyrAlaThrHisAspGlnGluGluAlaMetAlaIleSerAspGlnIleAla 205
  ::::: ::::: ::::: :::::
4247 TTGTTACTACGATCAAGAAAGAGCCCTTGCCCATGAGTGAATGGTTT 4296
206 ValMetLysAspGlyValIleGlnGlnIleGlyArgProLysGluLeuTyr 222
  ||| ::::: ::::: ||| ||| ||| :::::
4297 GTCATGATGAAGTGAATTTGTTCACTGCGGAACCACTGATATTTA 4346
222 rHisLysProAlaAsnGluPheValAlaThrPheIleGlyArgThrAsnI 239
  | ::::: ||| ::::: ::::: :::::
4347 TGATGAGCCCAATTAATCATTTTGTGCTAATTTATGGAATCTAATA 4396
239 IleIleProAlaAsnLeuGluLysArgSerAspGlyAlaTyrIleValPhe 255
  |||| ::::: :::::
4397 TTATTAACGTTACCATGATTGAA.....GACTATCTGTGCTCC 4434
256 SerAspGlyTyrAlaLeuArgMetProAlaLeuAspGlnValGluGlu 272
  ::::: ||| :::::
4435 TTTAACGGGAAGAAATTTGAATCTGAGCGGAGTGGCGCCCTAATGA 4484
272 nAlaIleHisValSerIleArgProGlu.....GluPheIleLysAspG 287
  : ::::: ||| ||| ||| |||
4485 GCCTGTGAAGTGGTATTCGCTCGAAGATCTTCAAAATTAATCTTCACG 4534
287 LuSerGlyAspIleGluGlyThrIleArgAspSerValTyrIleuGlyLeu 303
  ||::: ||| :::::

```







```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
  Quality: 583.50      Length: 367
  Ratio: 2.372         Gaps: 6
  Percent Similarity: 67.030   Percent Identity: 35.695

```

alignment block:

US-09-769-787-162 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

1 MetSerGluLeuLysIleIleAsnAlaLysLysIleTyr.....HisAs 15
|||||.....:|||||:|||||
2287093ATGGCTTGGATTGAGTCAGGCAACCGGCGTATCCCGGACGCA 2287044
15 pValProValIleGluAsnLeuAsnIleThrIleProLysGlySerLeuP 32
| |||.....:|||||:|||||
2287043CCGACCGCGCTGGATGCGCTCGACCTGATCGCGGATGGCGAGTTGC 2286994
32 hethrLeuLeuGlyAlaSerGlyCysGlyThrThrLeuLeuArgMet 48
:|||||:|||||:|||||:|||||
2286993TTGTCCGTGGGGCCGTCGCGATGTGGCAAGACGATGCTGATGCGGATG 2286944
49 IleAlaGlyPheAsnSerIleGluGlyGluPheThrPheAspPhe 65
:|||||:|||||:|||||:|||||
2286943GTGGCTGCTGGAGACGCTGAGCTGGGCGTATCCGATCGCGAGCG 2286894
65 rLysIleAsnAsnMetGluProSerLysArgAsnIleGlyMetValPheG 82
:|||||:|||||:|||||:|||||
2286893ACGATCGACGAGGTGATCCCAAGATCGTATGTCGATGCTGTTC 2286844
82 lnsnTyrrAlaIlePheProHisLeuThrValAlaArgAspAsnValAlaPhe 98
|||||:|||||:|||||:|||||
2286843AGAACTACGCTCTACCCGACATGACGGTGGCGAAGCATGGCTTC 2286794
99 GlyLeuMetGlnLysValProLysGluLeuIleGlnGlnThrAs 115
:|||||:|||||:|||||:|||||
2286793GCGTTGAAGTCCGCAAGATCGGCAAGCGGAGATCCGCGGGTGTCT 2286744
115 nLysTyrrLeuGluLeuMetGlnIleAlaGlnTyrrAlaAspArgLysProA 132
:|||||:|||||:|||||:|||||
2286743TCCGCGACGCAAAATTGCTGTATCTGCAATCTTATCTGATCGCAAGCGCA 2286694
132 sPlyLysLeuSerGlyGlyGlnGlnIleArgValThrLeuAlaCysAlaLeu 148
|||||:|||||:|||||:|||||
2286693AAGATCTCTCCGCGCGCAACGCGGAGGGGCGATGGCGTGCATTC 2286644
149 AlaValAsnProSerValIleLeuMetAspGluProLeuSerAsnLeuG 165
:|||||:|||||:|||||:|||||
2286643GTGGGGGCGCCACAGGTATTCGTGATGACGCAACCGCTGTCCAAATCTTGA 2286594
165 uAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisG 182
|||||:|||||:|||||:|||||
2286593CGCCAAACTCTCGCGGCAACCGCCATCGATCGCGCGGTACACAGGCG 2286544
182 luvAlaGlyIleThrThrValTyrrValThrHisAspGlnGluAlaMet 198
:|||||:|||||:|||||:|||||
2286543AACGCGGTACGACCGCTGATGTCACTACGACGACGAGTGCAGCGCATG 2286494
199 AlaIleSerAspGlnIleAlaValMetLysAspGlyValIleGlnGlnI 215
:|||||:|||||:|||||:|||||
2286493ACGATGGGTGACCGCTCGCGGTCTGTCTGACGCTGTCTGCTGCAACAGTG 2286444
215 eGlyArgProLysGluLeuTyrrHisLysProAlaAsnGluPheValAla 232
:|||||:|||||:|||||:|||||

```

```

2286443TCTCGCCTCGAGAGCTTACCGCAACCGGCGCAACGTGTCTGCGCG 2286394
232 hrhPheIleGlyArgThrAsnIleIleProAla...AsnLeuGluLysArg 247
|||||:|||||:|||||:|||||
2286393GCTTACGCTGGTTC.....CCGCGATGAACTGTTTCAAGCTT 2286356
248 SerAspGlyAlaTyrrIleValPheSerAspGlyTyrrAlaLeuArgMetP 264
|||:|||||:|||||:|||||
2286355TCCATCGCCCATTCACAGGTGTCTACGTGATGTCGAGATCTCTGCTGCC 2286306
264 oAlaLeuAspGlnValGluGlnAlaIleHisValSerIleArgProg 281
| |||.....:|||||:|||||
2286305GCGTGGCGTGTGAGCGGACCGGAGCTCATTCGATGCTGTGCGCCCG 2286256
281 luvGluPheIleLysAspGluSerGlyAspIleGluGlyThrIleArgAsp 297
|||||:|||||:|||||:|||||
2286255AACATTTGGAGCTGGCGGCGCGC..... 2286230
298 SerValTyrrLeuGlyLeuAsnThrAspTyrrPheIleGluThrGlyPheAl 314
:|||||:|||||:|||||:|||||
2286229.....ATCGAGATGACGTCGACATGTGTGCAAGAACTTGGACGCGA 2286189
314 aSerLysIleGln.....ValSerGluGluSerThrPheGluGluA 328
:|||||:|||||:|||||:|||||
2286188CGCTACTGTATGCGCGAATCGTGTGCGCGCGCTGCGCAATGACACAGT 2286139
328 spleu.....GlnLysGlyAsnArg 334
:|||||:|||||:|||||:|||||
2286138CAATCGTGTGAGTGGAGCGCGCGCGCGCGCGCGGATGAGTCGC 2286089
335 IleArgLeuArgIleAsnThrGlnLysLeuAsnIlePheSerAlaAspG 351
:|||||:|||||:|||||:|||||
2286088GTGCGGTATGTCCACACCGGAGACACTGCTTCTCCCGCTGACGCG 2286039
351 y 351
2286038g 2286038

```

seq\_name: /cgn2\_6/plodata/1/lna/6B\_COMB.seq:US-08-858-207A-52

seq\_documentation\_block:

; Sequence 52, Application US/08858207A

; Patent No. 6348328

; GENERAL INFORMATION:

; APPLICANT: Black, Michael

; APPLICANT: Hodgson, John

; APPLICANT: Knowles, David

; APPLICANT: Nicholas, Richard

; APPLICANT: Stodola, Robert

; TITLE OF INVENTION: No. 6348328e1 Compounds

; NUMBER OF SEQUENCES: 552

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: USA

; ZIP: 19406-0939

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/858,207A

; FILING DATE: 09-MAY-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: 60/017670

; FILING DATE: 14-MAY-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Gimm1, Edward R

; REGISTRATION NUMBER: 38,891

```
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1828 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-52

alignment_scores:
Quality: 442.00      Length: 90
Ratio: 4.966        Gaps: 0
Percent Similarity: 98.889      Percent Identity: 96.667

alignment_block:
US-09-769-787-162 x US-08-858-207A-52 ..

Align seg 1/1 to: US-08-858-207A-52 from: 1 to: 1828

274 ILEHISVALSERILEARGPROGLUGLUPHEILELYSASPGLUSERGLYAS 290
|||||
3 ATTCATGTAGAGCATTCGTCGCCAAGAGTTTATCAAGATGATCTGGAGA 52
|||||
290 PLEGLUGLYTHRILEARGASPERVALTYRLEUGLYLEUASNTHRAPPT 307
|||||
53 TATTAGAGCACTATTAGCGATAGCGCTCTATCTTGAGCTAAATAGAGAT 102
|||||
307 YRPHLEIGLITHRIGLYPHEALASERLYSILEGINVALSERGLUGLUSER 323
|||||
103 ATTTCATTGAGACAGGTTTGGCTCAAAAATTCAGTTAGTGAAGATCA 152
|||||
324 THRPHLEGLUGLUSPLEUGLINSGLYSANARGLILEARGLEUARGILEAS 340
|||||
153 ACTTTTGAAGAGAGACTACAAAAGCAATCTATTCTGTACGATCA 202
|||||
340 NTHRGLINSLEUASNILEPHERSERIALASPGLYSERGINASLEULEL 357
|||||
203 TACGCAAAATTTAAACATCTTTCTGCGAGATGTTCCCAAAACGTGATCA 252
|||||
357 YSGLYVALASNHSGLYTHR 363
|||||
253 AAGGAGTCACCATGGAAG 272
|||||

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-578-158-1

seq_documentation_block:
; Sequence 1, Application US/08578158
; Patent No. 5830690
; GENERAL INFORMATION:
; APPLICANT: Govrishankar, Jayaraman
; APPLICANT: Bhandari, Poonam
; APPLICANT: Rajkumari, Kaveti
; TITLE OF INVENTION: USE OF DNA ENCODING A PROMOTER OF
; TITLE OF INVENTION: PROMOTERS ALONG WITH CIS REGULATORY ELEMENTS AND
; TITLE OF INVENTION: A NOVEL PROCESS FOR PRODUCING POLYPEPTIDES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578, 158
FILING DATE: 29-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 07064/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-578-158-1

alignment_scores:
Quality: 418.00      Length: 227
Ratio: 2.580        Gaps: 1
Percent Similarity: 71.366      Percent Identity: 37.004

alignment_block:
US-09-769-787-162 x US-08-578-158-1 ..

Align seg 1/1 to: US-08-578-158-1 from: 1 to: 1260

19 ILEGLUASNLEUASNILETHRILEPROLYSGLYSERLEUPHERLEULE 35
|||||
452 GTAAAGACGCCAGCTCTGCCATTGAAGACGCCGAGATTTGTCATCAT 501
|||||
35 UGLYVALASERGLYCYSGLYLYSTHRLHEULEUARGMETLEALAIYP 52
|||||
502 GGGATTATTCGCCGCTGTAATCCACAAATGATGAGCGCTCTCATATGCC 551
|||||
52 HEASNSERILEGLUGLYGLUPHERYRPHASPTHRHLSLEASN 68
|||||
552 TGATTGAACCCACCCGCCGCAAGTGTGATTTGATGTTGATATTTGCC 601
|||||
69 ASNMET.....GLUPROSERLYSARGASNILEGLYME 79
|||||
602 AAATATCCGACGCCGAGCTCCGTGAGTGGCGCAAAAAGATTTCGAT 651
|||||
79 VALPHLEGLNASNTYRALAILEPHERPROHISLEUTHRYVALARGASPNV 96
|||||
652 GGTCTTCCAGTCTTTGGCTTAATGCCCATATGACCGTGTGACATCA 701
|||||
96 ALAIPHLEGLYLEUMETGLINSGLYSVALPROLYSGLUGLULEULEIN 112
|||||
702 CTGCGTTCGTATGGAATTTGGCCGAATTAATGCCGAAGACGCCGGAA 751
|||||
113 GLNTHRASNLYSTYRLEUGLULEUMETGLINLEALAGINTYRALASPAR 129
|||||
752 AAAGCCCTTGATGCACCTGCCAGTGGCGGTGGAATTTATGCCACAG 801
|||||
129 GLYSPROASPLYSLEUSERGLYGLYGLINGLNGLNARGVALTHRYLEUALAC 146
|||||
802 CTACCCGGATGACCTCTGCGGGGATGCGTCAACGTGTGGATGAGCC 851
|||||
146 YSALALEUALVALASNPROSERVALLEULEUMETASPGLUPTOLEUSER 162
|||||
852 GCGCGTTAGCATTAATCCGATATTAATTAATGAGAACCCCTTCG 901
|||||
163 ASNLEUGLUALALYSLEUARGLEUASPHEMARGINALALEARGLUIT 179
|||||
902 GCGCTCGATTCATTAATTCGACCGAGATGCGAGATGCGTGGAAAT 951
|||||
179 EGLNHSGLUVALGLYILETHRTHRVALTYRVALTHRHSASPLNGLUG 196
```



```
; Sequence 3, Application US/08919573
; Patent No. 6346392
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; TITLE OF INVENTION: ATP-BINDING PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,573
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50597
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SPOUNCE CHARACTERISTICS:
; LENGTH: 732 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-919-573-3

alignment_scores:
      Quality: 345.00      Length: 238
      Ratio: 2.091      Gaps: 3
Percent Similarity: 69.328      Percent Identity: 31.513

alignment_block:
US-09-769-787-162 x US-08-919-573-3 ..

Align seg 1/1 to: US-08-919-573-3 from: 1 to: 732

4      llelystlleasnalalyslysleryrlhsapvalprovalillegl 20
      |||||
16      atpaaatggaatattacatpaaatctttggaagaagatgacgattgcaa 65
      |||||
20      uasnleuasnillethrilleprolysglyserleupherleuenglya 37
      |||||
66      gggcatcactcgacgatttaaaagagagagctgtctatcattcgctc 115
      |||||
37      laserglycysglylysthrthrleuenaargmetilleaglypheasn 53
      |||||
116      cttcagcagagtggaatctacgttgcctctatgaatttggtgag 165
      |||||
54      serlleuglyglyluhphetyrphensapsphirlystle..... 67
      |||||
166      gaaccaaccaaaggaggttatcttgaaggagcgtatattacgacaa 215
      |||||
68      ....asnsmetcluproserlysarasnilleglymetvalpheglna 83
      |||||
216      gaagaattacctgtttggcattgcgtgagacaagatggcgttttccaac 265
```

```
83      snTyralallepheprohisleuthrvalargaspasnaalalephgly 99
      .....
266      aatttaacctcttcttctaatatgactgattgaaatattacctgttca 315
      .....
100      leuuetglnlyslsvalprolysglucluueuilegin...glntrnas 115
      .....
316      cctatcagaccacaaagtgaaagcaggaagttcctggaagacagacc 365
      .....
115      nlystlyrleugluueuetglnilleaglnTyralasparglyspcca 132
      |||||
366      aaaccttttgaaaaaagttggtttgccagatnaagcagacccttatccac 415
      .....
132      splsleuserglyglynglnlnargvalthrleualaCysAlaleu 148
      .....
416      agactctctcggtggcagacaaacagcattgccatcccggtgtg 465
      .....
149      Alalvalasnproservalleuueuetaaspqluproleuseranleugl 165
      .....
466      gctatggaaccagatgtyttgctcttgacgacacacacattacgcttga 515
      .....
165      uAlalysleuarlyleuaspmetarglnAlaileargluileglnhsg 182
      .....
516      tcctgaatagtgtagtgagctattggcgttatgcagaccttgcc...A 562
      .....
182      luvalglyilethrthrvaltyrvalthrhisaspglngluAlamet 198
      .....
563      agtcaggatgactatgattctgtaacacatgagatggattggccgt 612
      .....
199      Alalieserapnglnillealvalmetlyasppllyallieglnlnil 215
      .....
613      gagtggcagatcggttatcttattgacacacgctgtgtgagaga 662
      .....
215      eglYargprolysgluueuTyrlhslyslprolaasngluhpevala1at 232
      |||||
663      tgggacacactgacgagattttggaacaaacacagaaacagcgttaag 712
      .....
232      hrpheelleglyarg 236
      |||||
713      atttcttgagatpaa 726

seq_name: /cgn2_6/prodata/1/lna/6B_COMB.seq:US-08-919-573-1

seq_documentation_block:
; Sequence 1, Application US/08919573
; Patent No. 6346392
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; TITLE OF INVENTION: ATP-BINDING PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,573
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50597
```

```

TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 609-520-3214
:
: TELEFAX: 609-520-3259
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 735 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
:
: US-08-919-573-1

alignment_scores:
      Quality: 345.00      Length: 238
      Ratio: 2.091      Gaps: 3
      Percent Similarity: 69.328      Percent Identity: 31.513

alignment_block:
US-09-769-787-162 x US-08-919-573-1 ..

Align seg 1/1 to: US-08-919-573-1 from: 1 to: 735

4 11elysileleasnala1alysiletyrhisaspvalprovalilegl
|||||
16 ATAAATTTGAAATTTACATTAATCTTTGAAAGATGAGTATTTGAA
|||||
20 uasneusnilethrlleprolysglyserleuphetrlleu1ya 37
|||||
66 GGGCATCAACCTCGAGATTAAAGAGAGAACGTTGCTTATCATCGGTC
|||||
37 laSerGlyCysGlyThrThrleuLeuArgMetilealaglypheasn
|||||
116 CTTGAGGGAGTGGAAATCTACCTTGCTCTATGCAATTTGTTGAG
|||||
54 SerileguglygluPhetYrPheaspThrlyle..... 67
|||||
166 GAAGACCAAGGAGGAGGTATCTTTGAGGAGTCGATATTACGACAA
|||||
68 ...AsnsmetGluProSerlyshArgAsnileglmetvalpneqla
|||||
216 GAAGATGACCTGTTGCCATCGTGAGAAAGATGGCATGTTTCCAA
|||||
83 snTyralallepProhisleuThrValArgAspAsnValAlaphegly
|||||
266 AATTTAACCTCTTCCATATGACTGTGATGAAATATTTACCTGTCA
|||||
100 leumetGlnlysValProlysgluglu1leqln...GlnThrAs
|||||
316 CCTATCAAGACCAAGGTGAAGCAGGGAAGTTGCTGAGAACAGAGCCA
|||||
115 nlystYrleuGluLeuMetGlnlealaglnTyralAsparGlyspica
|||||
366 AAAGCTTTTGGAAAAAGTTGTTGCCAGATTAAGCAGACGCTTATCCAC
|||||
132 splyleuSerGlygluGlnGlnArgValThrleuAlaCysAla1leu
|||||
416 AGAGTCTGGGGTGAGGAGCAACAAGGATTCGCGGCGGTTG
|||||
149 AlaValAspSerValleuLeuMetAspGluProleSerAsnleu1
|||||
466 GGTATGGAACCAAGATGTTGCTTTGAGAACCAACTTCAAGCCTTGA
|||||
165 uAla1ysleuArgLeuAspMetArgGlnAla1leargGlu1leGlnHis
|||||
516 TCCTGAATGTAGGTGAGGTATGCTGTATGCAAGACCTTGCC...A
|||||
182 luValGlyleThrThrValTyValThrHisAspGlnlu1leuAlaMet
|||||
563 ACTCAGGGAGTACTAGTTATCGTAACCAATGAGATGGATTGGCCGT
|||||
199 Ala1leSerAspGln1leAla1ValMetlysAspGluVal1leGlnGln
|||||

```

```

613 GAGTGGCAGATCGTGTATCTTATGCGAGACGGTGTGTTGAAGA 662
215 eGlyArgProlysgluLeuTyRHisLysProAlaSnGluPheValAlaT 232
663 TGGACACCTGAGCAGATTTTGAACAAACCAAGAACAGGACTTAAG 712
232 hrPheileGlyArg 236
713 ATTCTCTGAGCTAAG 726

seq_name: /cgn2_6/plodata/1/ina/6B_COMB.seq:US-08-858-207A-49

seq_documentation_block:
: Sequence 49, Application US/08858207A
: Patent No. 6348328
:
: GENERAL INFORMATION:
:   APPLICANT: Black, Michael
:   APPLICANT: Hodgson, John
:   APPLICANT: Knowles, David
:   APPLICANT: Nicholas, Richard
:   APPLICANT: Stodola, Robert
:   TITLE OF INVENTION: No. 6348328el Compounds
:   NUMBER OF SEQUENCES: 552
:   CORRESPONDENCE ADDRESS:
:   ADDRESS: SmithKline Beecham Corporation
:   STREET: 709 Swedeland Road
:   CITY: King of Prussia
:   STATE: PA
:   COUNTRY: USA
:   ZIP: 19406-0939
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Diskette
:   COMPUTER: IBM Compatible
:   OPERATING SYSTEM: DOS
:   SOFTWARE: FastSeq for Windows Version 2.0
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/858,207A
:   FILING DATE: 09-MAY-1997
:   CLASSIFICATION: 435
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 60/017670
:   FILING DATE: 14-MAY-1996
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Gimmi, Edward R
:   REGISTRATION NUMBER: 38,891
:   REFERENCE/DOCKET NUMBER: P50475
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 610-270-4478
:   TELEFAX: 610-270-5090
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 49:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1460 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:
: US-08-858-207A-49

alignment_scores:
      Quality: 339.50      Length: 202
      Ratio: 2.219      Gaps: 3
      Percent Similarity: 75.743      Percent Identity: 34.653

alignment_block:
US-09-769-787-162 x US-08-858-207A-49 ..

Align seg 1/1 to: US-08-858-207A-49 from: 1 to: 1460

2 SerGluilelysileleasnala1alysiletyrhisaspvalprova
|||||
367 TCTATGTGAATTAAGAAATATCAATAAAGTCTTGGAGACAAACAAT
|||||

```

18 *llleGluAnleuAsnlllethrlleProlysglySerleuPhehrleu* 35  
417 *CCTGCTGATTTTCAGCTTAAGATTCTCGAAGAACAATCTGGCTATCG* 466  
35 *eugllyAlaserglycsglylsthrrhrleuAermeTleAlagly* 51  
467 *TTGGACCTTCCTGGTGAAGATAGAACACTCTTTTACAGTACTCTTGCGGGT* 516  
52 *PhasnSerlleGlnGlyGlyIubuheryrPhasnpsprhlyrIleas* 68  
517 *CTTGAACCACTTGATTCACAGGCAAAATCTTTATATGACACACCTTAGA* 566  
68 *nasnmetGluProSerLysArgAsn...lleGlyMetValPheGlnAsn* 84  
567 *GCTGAGATGAATTGCAGAACCCCAACCTACTGSGATTTGTCTCCAAAGATT* 616  
84 *ylrallePheProhIstleuthrValAlArgAspAsnValAlaPheglyLeu* 100  
617 *TTCAACTATTTCCTCATCTATCATAGTTGGAAAAATTTGCACCTTTATCGCCT* 666  
101 *MetGlnLysLys...ValProLysGlnLubuleuIleGlnGlnThrAsnly* 116  
667 *GTGAAGACCATGGGAATGAGAGAGAAAGGCTGAGAAAGAGCGAGACTGG* 716  
116 *stYrLeuGluLeuMetGlnIleAlaGlnTyAlaAspArgLysProAspL* 133  
717 *ACTCTTGGGAACAGTTTAGACTGGAGGACACGACAGACTCTATCTCTTCT* 766  
133 *ysLeuSerGlyGlnGlnGlnGlnArgAlThrAlaIacyAlaLeuAla* 149  
767 *CACATATCTGTGGGCAAAAGCAGCGGGTGGCTTTGGCGCGGTATATG* 816  
150 *ValAsnProSerValleuLeuMetAspLubProLeuSerAsnLeuGln* 166  
817 *ATTGACCCCAAGAATCATTTGGCTACACTGAACAACACTTCCGCCCTGCATCC* 866  
166 *alysLeuArgLeuAspMetArgGlnAlaIleArgGlnIleGlnHISGlnV* 183  
867 *AGATATTAACGTTTGGAGTGTGGAACACTATATCTTGCA...AATGGGAAC* 913  
183 *alIlylleThrThrValTyAlaThrIstAspGlnGlnAlaIaMetAla* 199  
914 *TTTGGAGTGAACACAGATTGTGGTTACCATCGATTTTGCAGTTTGGTGAAT* 963  
200 *lleSer* 201  
964 *ATCGCC* 969



```

509 GTTCCAGCCAAAGCAGCAATATAGCGATGCTTCCAGAAATTACGCCTT 460
86 ephProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnL 103
   ::::::::::::::::::::
459 GTATCCGATATGACGGTTTACGACAAATGCGCTTGTCTGCAAGATGC 410
103 yslYsValProLysGluGluLeuIleGlnGlnThrAsnLysTyrLeuGlu 119
   ::::::::::::::::::::
409 AAAAATCTGCCAAAGAGCTGATTTGATGAGCGGCTCAACTGGCGCGCA 360
120 LeuMetGlnIleAlaGlnTyrAlaAspArgLysProAspLysLeuSerG 136
   ::::::::::::::::::::
359 ATTCTGCGCTTGGCTGAGTACCTGAACGATAAGCGGGGCGCTTCCGG 310
136 yglYglnGlnGlnThrValThrLeuAlaCysAlaLeuAlaValAsnPro 153
   ::::::::::::::::::::
309 CGGGCAACGTCAAGCGAGTGGCGCTTGGCGGCGCATTTACGCGAAGCG 260
153 eValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArg 169
   ::::::::::::::::::::
259 GCGTGTATTATGATGATGACCGCTCTTAACTGATGACCAAGCTCGCG 210
170 LeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValGlyIleTh 186
   ::::::::::::::::::::
209 GTGCAATGCGCGCAGAGATCAGCAGCTGCATCGAAGAACTGAAACACCA 160
186 rThValTyrValThrHisAspGlnGluGluAlaMetAlaIleSerAsp 203
   ::::::::::::::::::::
159 CATGATCTACGTGACCCAGATCAGACCGAAGCGATGACATGGCGAGCG 110
203 InIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 219
   ::::::::::::::::::::
109 GGATGTGATATATGAAAGACGCGGATGTTTCAGCAAGTAGTGCGCGCAA 60
220 GluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGly 235
   ::::::::::::::::::::
59 ACCGTTTATTAACCAACCGCGCAATATGTTGTTCCGAGATTATTTGA 12

```

seq\_name: gb\_est2:BG855631

seq\_documentation\_block:

LOCUS BG855631 720 bp mRNA linear EST 29-MAY-2001

DEFINITION 1024043612.x2 C. reinhardtii CC-1690, normalized, lambda zap II

ACCESSION Chlamydomonas reinhardtii cDNA, mRNA sequence.

VERSION BG855631

KEYWORDS BG855631.1 GI:14236815

SOURCE EST.

ORGANISM Chlamydomonas reinhardtii.

REFERENCE Chlamydomonas reinhardtii

AUTHORS Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

1 (bases 1 to 720)

Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,

McDermott, J. P., Sillflow, C., Stern, D. and Surzycki, R.,

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants; Project phase 2

Journal, unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu

FEATURES

1. /720

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/gb\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, lambda zap

II"

/note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:

XhoI. This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 157 a 215 c 180 g 166 t 2 others

ORIGIN

alignment\_scores: Quality: 468.50 Length: 235  
Ratio: 2.617 Gaps: 3  
Percent Similarity: 76.170 Percent Identity: 41.702

alignment\_block:

US-09-769-787-162 x BG855631/rev ..

Align seg 1/1 to reverse of: BG855631 from: 1 to: 720

```

63 AspAspThrLysIleAsnAsnMetGluProSerLysArgAsnIleGlyme 79
   ||| ::::::::::::::::::::
710 GATGGCGTCGATNTGTCACAGATTCCGCCCTTACCTCGCGCCCATCATAT 661
79 tValPheGlnAsnTyrAlaIlePheProHisLeuThrValArgAspAsn 96
   ::::::::::::::::::::
660 GATGTTCTCAGTCTTACGCGCTGTTCCCATATATGACCTGGAAACAGAA 611
96 aAlaIlePheGlyLeuMetGlnLysLysValProLysGluGluLeuIleG 112
   ::::::::::::::::::::
610 TCGCTTTTGGCTCGTAACAGACAACTACCGAAGCCGCAATTTGCCAGC 561
113 GlnThrAsnLysTyrLeuGluLeuMetGlnIleAlaGlnTyrAlaAsp 129
   ::::::::::::::::::::
560 CGGGTCATGATGATGCTCGCGCTGTCATATGACAGAGTTCCGCAACG 511
129 gLysProAspLysLeuSerGlyGlyGlnGlnGlnThrValThrLeuAla 146
   ::::::::::::::::::::
510 CAAACCGCATGACCTTCCGCTGTCACAGCAACGTTGCGCCCTCGGCC 461
146 yAlaIleLeuAlaValAsnProSerValLeuLeuMetAspGluProLeu 162
   ::::::::::::::::::::
460 GAAGCTTTCGAGAGCGCCGAACTATTACTGCTCGATGACCGCATGGGC 411
163 AsnLeuGluAlaLysLeuArgLeuAspMetArgGlnAlaIleArgGlu 179
   ::::::::::::::::::::
410 GCGCTCGATTAAGAAAGCTCCGCGACAGATGACAGCTTGAAGCGTGAT 361
179 eGlnHisGluValGlyIleThrThrValTyrValThrHisAspGlnGlu 196
   ::::::::::::::::::::
360 TCTGAGCGCGCTCGCTGATGCTGTGTGATGATGATGATGATGATGAT 311
196 lValAlaMetAlaIleSerAspGlnIleAlaValMetLysAspGlyVal 212
   ::::::::::::::::::::
310 AGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261
213 GlnGlnIleGlyArgProLysGluLeuTyrHisLysProAlaAsnGlu 229
   ::::::::::::::::::::
260 GTCCAGATTTGGCGAAGCGAAGATGATGATGATGATGATGATGATGAT 211
229 eValAlaThrPheIleGlyArgThrAsnIleIleProAlaAsnGlu 245
   ::::::::::::::::::::
210 TAGCGCGGATTTATTTGCTGCTGATTAACGCTTTGAAGCGCTACTCAA 161
246 LysArgSerAspGlyAlaTyrIleValPheSerAspGly.....TyrAl 260
   ::::::::::::::::::::
160 GAGCGTCAGGAAGATGCGCTGCTGATTCGCGCGGCGCTGTCATCC 111

```



```

260 ateuatgmetproala...leuasplnialugluginlaiailehisv 276
    * |||||::||| ::|||::|||::|||::|||
110 ACTGAAGTCGACGCCGATGCCTCGGTGCTATACGTCGCCGGTACATG 61
276 alserllearproglugluhpheilleayaspolsuserglyaspiiegju 292
    ||::|||::|||::|||::|||::|||::|||::|||::|||::|||
60 TGGCGCTGCGCCCGGAAAAAATCATCATTGGAGAAGAGCGCCGCCCAAT 11
293 gjy 293
    |||
10 Gcr 8
seq_name: gb_gss:BH144062

seq_documentation_block:
LOCUS BH144062 618 bp DNA linear GSS 16-AUG-2001
DEFINITION TBEGS07H cTOG lycopersicon esculentum genomic clone cTOG28P11, DNA
sequence.
ACCESSION BH144062
VERSION BH144062.1 GI:15198524
KEYWORDS GSS.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyte; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 618)
AUTHORS van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and
Tankisley,S.
TITLE Tomato Demethylated Genomic DNA sequences
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
tomato demethylated genomic DNA
Insert Length: 1270 Std Error: 0.00
Seq primer: M13F-R
Class: shotgun.
FEATURES
Source Location/Qualifiers
1..618
/organism="Lycopersicon esculentum"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone_xref="cTOG28P11"
/clone_lib="cTOG"
/issue_type="Young leaves"
/dev_stage="12-14 weeks post harvest"
/lab_host="E.coli JM109"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; This library was made from short EcoRI digested
fragments of the genome of Lycopersicon esculentum ligated
into pBS (SK-). The fragments were cloned into the
methylation restrictive E.coli strain JM109 with the
purpose of enriching the library for non-methylated DNA
fragments. This procedure may enrich the pool of cloned
fragments in JM109 cells for sequences representing
expressed genes. Average insert size 1.27 kb."
BASE COUNT 124 a 204 c 170 g 120 t
ORIGIN
alignment_scores:
Quality: 425.50 Length: 201
Ratio: 2.875 Gaps: 4
Percent Similarity: 73.632 Percent Identity: 45.274
alignment_block:
US-09-769-787-162 x BH144062 ..

```

[illegible]

```

MEDLINE      21376150
COMMENT      Contact: Wing RA
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: rwing@clemson.edu
              Class: BAC ends
              High quality sequence stop: 809.
              Location/Qualifiers
FEATURES
  source      1. 874
              /organism="Bradyrhizobium japonicum"
              /strain="USD110"
              /db_xref="taxon:375"
              /clone_lib="B. japonicum BAC library"
              /lab_host="E. coli"
              /note="vector: pindigo536. Site_1: HindIII"
BASE COUNT   146 a      268 c      281 g      175 t      4 others
ORIGIN
alignment_scores:
  Quality:    417.00      Length:    272
  Ratio:      2.206      Gaps:      6
  Percent Similarity: 69.485      Percent Identity: 36.765
alignment_block:
  US-09-769-787-162 x AZ934163/rev ..
Align seg 1/1 to reverse of: AZ934163 from: 1 to: 874
41  G1yLysThrThrLeuLeuArgMetIleAlaG1yPheAsnSerIleGluG1 57
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 GGGCAAAACCCACGCGATGATAGTGGCTGGTTGAGTCCGACCCATG 811
    57 yG1yGluPheTyrPheAspThrLysIleAsnMetGluProSerL 74
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
810 GGAAGATTTTTCGCAAGACGCCGCTTAACAAATCCCGCCGACAA 761
    74 ySArGAsnIleG1yMetValPheGlnAsnTyrAlaIlePheProHISLeu 90
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
760 AGCGGCGACATGCGATGTTGTCAGATTATGCTTTGTGTCGCGACATG 711
    91 ThrValArGAspAsnValAlaPhe.G1yLeuMetGlnLysValProL 107
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
710 CGATGCCGGAATAATATNCTTCCCGCTC...GTCGCCGAATATCACA 664
    107 ySg1yGluLeuIleGlnGlnThrAsnLysTyrLeuGluLeuMetGlnIle 123
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
663 AGCCGNAAGCGCAGAACGCCGTCAGNCGCGCTCCCGCATGATCAGATG 614
    124 AlAGlnTyrAlaAspArgLysProAspLysLeuSerG1yGlnGlnG1 140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 AAACCCCTGGCGACCGCGCGCGCGGAGCTGTCGCGGTGACAGCA 564
    140 nArGValThrLeuAlaCysAlaLeuAlaValAsnProSerValLeuLeu 157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
563 GCGCGTGGCGCTCGCGCGCGCGCTTTCATCCGCAATCGCATCGTGTA 514
    157 eLAspG1yProlLeuSerAsnLeuG1yAlaLysLeuArGLeuAspMetArG 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
513 TGGAGAGGCCCTTGGCGCTGACAAAGCGCTGCGGAGAGATGCCAA 464
    174 G1AlaIleArGArgGluIleGlnHISG1yAlaG1yIleThrValTyrVa 190
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 CTGGAATCAAGCAACTGCACGAGACGATGGCGCTCGCTACGT 414
    190 lThrHISAspGlnG1yAlaLametalAlaIleSerAspGlnIleAlaVal 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 CACCGACGATCAGATGAACGCGCTCACCATGTCGACCGCATGGCGGT 364
    207 eLysAspG1yValIleGlnGlnIleG1yArGProLysGluLeuTyrHIS 223

```

```

363 TCAACGACGCGATCTGCACAGATGCACAGGCCCGACGCGCTGTATGAG 314
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 LysProAlaAsnGluPheValAlaThrPheIleG1yArGThrAsnIleL 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 CATCCGGTGAACAGCTTCGTGCGCTCACTTCATCGCGCAGAACATATGCT 264
    240 eProAlaAsnLeuGluLys...ArgSerAspG1yAlaTyrIleValPheS 256
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 GCGCGGACCGCTGACGAGCGGTGCAGAACGACTATTGCCGCGCGCTGG 214
    256 eArSpG1yTyrAlaLeuArGMetProAlaLeuAspGlnValG1yGluGln 272
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
213 CGGCGCGCGCGCGCGCTACCGCACAGCGGGGTCAATGATCAGCGCGGCG 164
    273 AlA...IleHISValSerIleArGProGluGluPhe...IleLysAspG1 287
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 GCATCGACCTCCCTGTCGTGCGCGCGGACGCGTCCGCTTGTCCCGGA 114
    287 uSerG1yAspIleGluG1y.....ThrIleArGAspS 298
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 CGGACCTCCAGCGGAAGACCGAACCGCGTGCAGCGGATGTCAGACAGA 64
    298 eArValTyrLeuGly 302
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 CCATCTATCTCTCGGC 50
seq_name: gb_est1:AV594537
seq_documentation_block:
LOCUS      AV594537      454 bp      mRNA      linear      EST 27-NOV-2001
DEFINITION AV594537 Bos taurus cartilage fetus Bos taurus cDNA clone
ACCESSION  E1CA014C12 3', mRNA sequence.
VERSION    AV594537
KEYWORDS   AV594537.1 GI:9710934
SOURCE     EST.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 454)
            Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
            Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
            Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL    21570554
MEDLINE    Contact: Yoshikazu Sugimoto
COMMENT    Animal Genetics Division
            Shriakawa Institute of Animal Genetics
            Odokura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES
  source      1. 454
              /organism="Bos taurus"
              /db_xref="taxon:9913"
              /clone="E1CA014C12"
              /clone_lib="Bos taurus cartilage fetus"
              /tissue_type="cartilage"
              /dev_stage="fetus"
              /lab_host="DH10B"
              /note="vector: pZL1, Site_1: SalI; Site_2: NotI; Poly A
              was deleted from a NotI site"
BASE COUNT   98 a      131 c      124 g      101 t
ORIGIN

```

alignment\_scores:                   Quality: 404.00                   Length: 151  
                                       Ratio: 3.311                   Gaps: 0  
                                       Percent Similarity: 80.795       Percent Identity: 50.331

## alignment\_block:

US-09-769-787-162 x AV594537/rev ..

Align seg 1/1 to reverse of: AV594537 from: 1 to: 454

```

77  TLEGLYMETVALPHEGLINASTYRVALAIEPHEPROHISLEUTHVALAR 93
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453  GTTGGATGTGTTTCAGTCTTACGGCTCTATCCACCCTGCAGTAGC 404
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93  GASPARVALALAPHEGLYMETGLINUSLYSVALPROLYSGLUGLUL 110
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403  AGAAGACATGCTATTGGCTGAACCTGGCTGGCGCAAAAAGAGGTGA 354
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110  euileglnlnhrasnlystyrluengluleumetglnilealaglntyr 126
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353  TTAAACCAAGCGTTAACAGAGTGGCGAAGTGTACACAGCGGCATTTC 304
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127  ALASPARGLYSPROASPLYSLSEUSERGLYGLINGLNGINLARGVALTH 143
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303  CTGGATCGCAAAACCGAAGCGCTCTCCGGGTGTCAGCGTCAAGGTGGC 254
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143  rleuAlaCysAlaLeuAlaValAsnProSerValleuLeuMetAspGlu 160
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253  GATTGGCCGTACGCTGTGGCGCGCAACCGATTATTTTCTCGATGAAC 204
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160  roLeuSerAsnLeuGlnAlaLysLeuArgLeuAspMetArgGlnAlaIle 176
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203  CGCTCTCCAAACCTCGATGCTGCACCTGGTGTGCAATGCTATCGAATC 154
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177  ArgGlnIleGlnIleGlnIleValGlyIleThrThValTyrValThrHis 193
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153  TCCCGCTGCTGATTAACGCCCTGGCGCGCAATGATTTACTCACCACGA 104
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193  pGlnGlnGlnAlaMetAlaIleSerAspGlnIleAlaValMetLysAsp 210
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103  TCAGGTGCAAGCGATGACGCTGGCCGCAAAATCGTGGTCTGGACCGG 54
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210  lYalIleGlnIleGlnIleGlyArgProLysGlnIleUeUryHisLysPro 226
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53  GTGGCGGCGCGCAGGTTGGGAAACCGCTGAGCTGATCACTATCCGCA 4
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227  Asn 227
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3  GAC 1
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: gb\_gss:AZ935459

seq\_documentation\_block:

LOCUS                   AZ935459                   545 bp       DNA       linear       GSS 24-Apr-2001  
 DEFINITION    BJ\_Ba000H18r B. japonicum BAC library Bradyrhizobium japonicum  
                   genomic, DNA sequence.  
 ACCESSION    AZ935459  
 VERSION      AZ935459.1   GI:13778216  
 KEYWORDS     GSS.  
 ORGANISM     Bradyrhizobium japonicum.  
               Bradyrhizobium japonicum.  
               Bacteria: Proteobacteria: alpha subdivision, Rhizobiaceae group:  
               Bradyrhizobium group: Bradyrhizobium.  
 REFERENCE    1 (bases 1 to 545)  
 AUTHORS      Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goloechea  
               ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.  
 TITLE        A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
               genome  
 JOURNAL      Genome Res. 11 (8), 1434-1440 (2001)  
 MEDLINE      21376150  
 COMMENT      Contact: Wing RA  
               Clemson University Genomics Institute  
               Clemson University

100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Class: BAC ends  
 High quality sequence stop: 502.  
 Location/Qualifiers

## FEATURES

## source

1..545  
 /organism="Bradyrhizobium japonicum"  
 /strain="USD110"  
 /db\_xref="taxon:375"  
 /clone\_lib="B. japonicum BAC library"  
 /lab\_host="E. coli"  
 /note="Vector: pIndigo536; Site\_1: HindIII"  
 BASE COUNT   86 a   167 c   197 g   93 t   2 others  
 ORIGIN

alignment\_scores:                   Quality: 385.00                   Length: 164  
                                       Ratio: 3.080                   Gaps: 0  
                                       Percent Similarity: 76.220       Percent Identity: 46.341

## alignment\_block:

US-09-769-787-162 x AZ935459/rev ..

Align seg 1/1 to reverse of: AZ935459 from: 1 to: 545

```

67  TLaSAsnMetGluProSerLysArgAsnIleGlyMetValPheGlnAs 83
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492  GTCACTTACATGCCCGCCGCAAGCGCGACATCGCGCTGATTCAGAA 443
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
83  nYrAlaIlePheProHisLeuThrValArgAspAsnValAlaPheGlyL 100
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442  CTACCGCGCTGTCCCGCATATGAGCGTCGGCGAAGCGTGTTCCTCC 393
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100  euMetGlnLysValProLysGlnIleUeUleGlnIlnTrAsnLys 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392  TGGCGCGCGCCCACTGCCGAAGCGAGCGCGCGAACAAGTGGCGCG 343
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117  TyrLeuGlnIleUeUetGlnIleAlaGlnTyrAlaAspArgLysPro 133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342  GCCCTCGCATGCTGCGGCTTGGCGGTACGAGGAACGCGCATCCGCA 293
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133  sLeuSerGlyGlyGlnGlnIlnArgValIThrLeuAlaCysAlaLeuAl 150
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292  GCTCTCCGGCGCCAGCGCCAGCGCGTGGCGCACGCGCATATGATCT 243
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150  alAsnProSerValleuLeuMetAspGluProLeuSerAsnLeuGlnAl 166
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242  TCAGAGCGCGCTGTGATCTGATGAGACGCGCATATCCGCGCTCGACA 193
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167  LysLeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnIleGln 183
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192  CACGTGGCGCAATCATGACATGACCTGCGCGCGTGCACCGCGCAT 143
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183  lGlyIleThrThValTyrValIThrHisAspGlnIleGlnIleAlaMet 200
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142  CGGCGCACCATCATCTACATCCACCATGATGATGAGCGGAGCGCTGAC 93
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200  leSerAspGlnIleAlaValMetLysAspGlyAlaIleGlnIlnIleGly 216
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92  TGAGCGACCGCGTCCGCTGATGAGACGCGGAGATCCAGATCGAC 43
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217  ArgProLysGlnIleUeUryHisLysProAlaAsnGluPheVal 230
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42  GAGCGGCGACGTCTGCACGACATCCCGCGCATCTCTTCGTC 1
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq\_name: gb\_gss:A0012091

seq\_documentation\_block:

LOCUS                   A0012091                   670 bp       DNA       linear       GSS 04-JUN-1998  
 DEFINITION    531Pg077070297 Cosmid library of chromosome II Rhodobacter

```

ACCESSION   spherae001
VERSION     A0012091
KEYWORDS    A0012091.1 GI:3177046
SOURCE      GSS.
ORGANISM    Rhodobacter sphaeroides.
            Rhodobacter sphaeroides.
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodospirillum rubrum.
REFERENCE   1 (bases 1 to 670)
AUTHORS     Choudhary, M., Mackenzie, C., Mouncey, N., Weinstock, G.M. and Kaplan,
            R.S.
TITLE       RasDB, the Rhodobacter sphaeroides Genome Database
JOURNAL     Unpublished (1998)
COMMENT     Contact: Choudhary, M.
            Department of Microbiology and Molecular Genetics
            University of Texas Medical School
            6431 Fannin Street, Houston, TX 77030, USA
            Tel: 713 500 5437
            Fax: 713 500 5499
            Email: madhuetung.med.utx.tmc.edu
            Seq primer: PBLuescript T7
            Class: shotgun.
FEATURES
Source      Location/Qualifiers
            1..670
            /organism="Rhodobacter sphaeroides"
            /strain="2.4.1T"
            /db_xref="taxon:1063"
            /clone="531PG0770297"
            /clone_1lb="Cosmid library of chromosome II"
            /lab_host="E. coli S17-1"
            /note="Vector: PLA2917"
BASE COUNT   120 a      232 c      203 g      115 t
ORIGIN
alignment_scores:
Quality:     377.00      Length:      225
Ratio:       2.299      Gaps:      4
Percent Similarity: 72.889      Percent Identity: 38.222
alignment block:
US-09-769-162 x A0012091 ..
Align seg 1/1 to: A0012091 from: 1 to: 670
5 LysIleIleAsnAlaLysIleTyRHIsaspValProValIleGluAs 21
|||||::: ||| ||| :::: |||:::
7 AAGATCGTG.....AAGCGATATGGCGGATGAGGCCATCCACGG 47
7 nleuAsnIleThrIleProlyGlySerIleuPheThrIleuGlyAla 38
|||||::: :::: |||::: :::: |||
48 CCGTCATCTCGAGTGGAGGACGGCGACTTCGTGGCCCTCGTGGGCC 97
38 eGlyCYsgIyLysThrThrIleuAlaGlyMetIleAlaGlyPheAsn 54
|||||::: |||::: |||::: |||::: |||::: |||::: |||
98 CGGGCTGGCGGAATCCACATCGCTCGCGATGATTTGGCGGCTGAGG 147
55 IlegIuGlyGluPheTyRPhaAspAptThrIyIleAsnAsnMetG 71
|||||::: :::: ||| :::: ||| :::: ||| :::: |||
148 ATCAGCGGGCGCCATATGCGATCGCGACGGCGTGGAGACGATATCA 197
71 uProSerIyAsnAsnIlegIyMetValPheGlnAsnTyRAlaIlePhe 88
|||||::: |||::: |||::: |||::: |||::: |||::: |||
198 GCCCAAGGGCCGCGACGTGGCGATGTCTTCAGATTACCGCGCTATC 247
88 rOhIleuThrValArGAspAsnValAlaPheGlyIleMetGlnLys 104
|||||::: |||::: |||::: |||::: |||::: |||::: |||
248 GCGCATATGACGCTGGCGCAACATCGGCTTCGGGCTTAATATCGCG 297
105 ValProLyGluGluIleuIleGlnIlnThraSnIyTyRLeuGluLeu 121
|||||::: :::: ||| :::: ||| :::: ||| :::: |||
298 GAGCCCGCGAGACGATTCGMAAGAACTCAGAGAAAGCCGCGCATCT 347

```

```

121 tginiiLeaGIntYrAlaSPaRySPRoAPlYSLeuSerGlyG 138
138 InGInGInaGValThrLeuAlaCysAlaLeuAlaValaSPRoSerVal 154
154 LeuLeuMetAspGluProLeuSerAsnLeuGluAlaLeuLeuArgLeuAs 171
171 pMetArgGInAlaIleArgGluIleGInnISGLValGlyIleThrThrv 188
188 AAlTValThrHisAspGInGluAlaMetAlaIleSerAspGInI 204
204 eAlaVal..MetLysAspGlyValIleGInGInIleGlyArgProLysG 220
220 IuLeu.TyrHisLysProAla 226
647 AACCTCTACAAACGGCCCTCC 667
seq_name: gb_gss:CNS07HFA

seq_documentation_block:
LOCUS      CNS07HFA                      557 bp    DNA       linear   GSS 02-OCT-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 26P08 of library NotreDamel
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION  AL610984
VERSION    AL610984.1 GI:15917169
KEYWORDS
SOURCE     African malaria mosquito.
            Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidae; Anopheles.
REFERENCE  1 (bases 1 to 557)
            Genoscope.
            Direct Submission
            Submitted (01-OCT-2001) Genoscope - Centre National de Sequenace :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
2 (bases 1 to 557)
            Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H.
            Direct Submission
            Submitted (01-OCT-2001) BWHI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT    This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.

FEATURES             Location/Qualifiers
     source           1..557
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="26P08"
                     /clone_lib="NotreDamel"
                     /note="end : SP6"
BASE COUNT          106 a      147 c      175 g      129 t
ORIGIN
alignment_scores:
Quality: 371.50      Length: 161
Ratio: 2.972        Gaps: 2

```

Percent Similarity: 77.640 Percent Identity: 47.826

Alignment block:

US-09-769-787-162 x CNS07HFA

Align seg 1/1 to: CNS07HFA from: 1 to: 557

```
118 LeuGIuLeuMetGlnIleAlaGlnIleAlaAspArgLysProAspLysLe 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 CAGGAGGCGTGAAGCTTCCGGTTACGGCATTCCGAACACGATGAACT 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 uSerGlyGlnGlnGlnIleAlaValThrLeuAlaCysAlaLeuAlaValA 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112 GTCTGGCGGTACAGACAGCAGCTGTCGCGTGGCGCGCGGTGATGCC 161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 sPProSerValLeuLeuMetAspGluProLeuSerLysLeuAlaLys 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 GACCCAAAGTCTGCTGTGATGAGCGCTGCTGCGGTGATTAATAAT 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 LeuArgLeuAspMetArgGlnAlaIleArgGluIleGlnHisGluValG 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 CAGCGTCTCATGATGAGCTTGAAGATTCACGATCAACGACGCTTGG 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 yIleThrThrValTyrValThrHisAspGlnGlnGlnAlaMetAlaL 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 CGTACACACCGCTCTTCGTGACACACGATGAGGCGCAAGCGGTGACAT 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 eAspGlnIleAlaValMetLysAspGlyValIleGlnGlnIleGlyArg 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
312 GCGATCGCTGGTGGTGGATGCGGCGACATGTACGCCAGATCGGCACG 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 ProLysGluLeuTyrHisLysProAlaAsnGluPheValAlaThrPhe 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
362 CCGGATGAAATTTATTCGTCGTCGCGCAGATCCCTTTGCGCGGTTTGT 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
234 eGlyArgThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGly 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
412 AGCGATGTGAATATTTCGCGGTCGTACGTCACGTCGC...GATAGGG 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 lAtyTrIleValPheSerAspGlyTyrAlaLeuArgMetProAlaLeu 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
459 CGGCGATGTGATCTGGCGGTAACTGTACGCTGACGCGCAGCG... 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
268 GlnValGluGluAlaIleHisValSerIle 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503 .....TAACTGTCCATGCTCGAT 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

seq\_name: gb\_gss:BH614583

seq\_documentation\_block:

LOCUS BH614583

DEFINITION 28BED9 Subclones from overlapping BAC clones spanning the hrp

cluster of *Erwinia carotovora* subsp. *atroseptica* *Pectobacterium*

*carotovorum* subsp. *atrosepticum* genomic, DNA sequence.

ACCESSION BH614583

VERSION BH614583.1

KEYWORDS GSS.

SOURCE *Pectobacterium carotovorum* subsp. *atrosepticum*.

ORGANISM *Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium*.

REFERENCE 1 (bases 1 to 569)

AUTHORS Bell, K.S., Avrova, A.O., Holeva, M.C., Cardie, L., Morris, W., De Jong

TITLE 'W', Toth, I.K., Waugh, R., Bryan, G.J. and Birch, P.R.J.

JOURNAL Sample sequencing of a selected region of the genome of *Erwinia*

carotovora subsp. *atroseptica* reveal a candidate phytopathogenicity

genes and allows comparison with *Escherichia coli*

Unpublished (2002)

Contact: Bell KS

Scottish Crop Research Institute

Invergowrie, Dundee, United Kingdom, DD2 5DA

Email: kbells@scri.sari.ac.uk

Class: BAC subclone.

FEATURES

source

Location/Qualifiers

1..569

/organism="Pectobacterium carotovorum subsp. atrosepticum"

/strain="SCRI039"

/db\_xref="taxon:29471"

/clone\_lib="Subclones from overlapping BAC clones spanning

the hrp cluster of *Erwinia carotovora* subsp. *atroseptica*"

/note="Vector: pGEM 3zf; DNA from BAC clones Eca2B8 and

Eca1C22 was nebulised and cloned into pGEM 3zf and

sequenced with SP6 or T7 primers"

BASE COUNT 131 a 159 c 155 g 124 t

ORIGIN

alignment\_scores:

Quality: 365.50 Length: 158

Ratio: 2.948 Gaps: 1

Percent Similarity: 78.481 Percent Identity: 45.570

alignment\_block:

US-09-769-787-162 x BH614583

Align seg 1/1 to: BH614583 from: 1 to: 569

```
8 AsnAlaLysLysIleTyrHis...AspValProValIleGlnAsnLeuAs 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 AACGTCACAGAACCTACGACGCGGACCGCTGATGTCACAAACCTGAA 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
23 nIleThrIleProLysGlySerLeuPheThrLeuGlnGlyAlaSerGlyC 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 TCTGATATTTCGGAAGGTGAATTCTGACCCGCTTGGGCGCGCGCT 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 ySgLyLysThrThrLeuLeuArgMetIleAlaGlyPheAsnSerIleGlu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
196 CGGGAAGAAACACACAGTTGATGATCTCGCGGCTTTGAACCCACAG 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 GlyGlyGluPheTyrPheAspThrLysIleAsnAsnMetGluProse 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 CAGGAGAAATCTGCTGCGCATGCGACCGTTGCATCATCTCCGCCGCA 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 rLysAlaLysIleGlyMetValPheGlnAsnTyrAlaIlePheProHis 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 TCAGCGCGACATCGCATGTTGTTTCAAGACTACGACGATTTCCACACA 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90 eUThrValArgAspAsnValAlaPheGlyLeuMetClnLysLysValPro 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
346 TCACTGTGGCGGAGAAATCTGCGCTTCCGCTGTCATCCGCGCTTAAAC 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 LysGluGluLeuIleGlnGlnIlePheAsnLysTyrLeuGluLeuMetGln 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 CGGTCGATATCAAGAGAAAGTCGATCGCTGCTGATCGCTTAAGCT 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 eAlaGlnTyrAlaAspArgLysProAspLysLeuSerGlyGlnGlnGln 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 GACTACCTAGCGGATCGGATACCCCGACAGATGTCGCGGCGGACACAG 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 lInArgValThrLeuAlaCysAlaLeuAlaValAsnProSerValLeuLeu 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
496 AGGCTGTGGCGGCTGCTCCGCGCTGCTGCAACCGAAGATGGTGTG 545
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 MetAspGluProLeuSerAsnLeu 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
546 ATGGATGAACCACTGGGTGCGCTG 569
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

seq\_name: gb\_gss:U80962

seq\_documentation\_block:

LOCUS U80962

DEFINITION U80962 Pseudomonas aeruginosa PA01 (ATCC15692) Pseudomonas

aeruginosa genomic, DNA sequence.

ACCESSION U80962

VERSION U80962.1

KEYWORDS GSS.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Pseudomonas aeruginosa.	Pseudomonas aeruginosa.	1 (bases 1 to 804)	Gensberg, K., Lowe, A.M. and Smith, A.W.	Genbank accession PA01 partial sequence of DNA which hybridised with a degenerate oligonucleotide probe based on the N-terminal amino acid sequence of Pseudomonas aeruginosa OprG	Unpublished (1996)	
Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.					Contact: Gensberg K	
FEATURES						
source	Location/Qualifiers					
	1..804					
	/organism="Pseudomonas aeruginosa"					
	/strain="PA01 (ATCC15692)"					
	/db_xref="taxon:287"					
	/clone_lib="Pseudomonas aeruginosa PA01 (ATCC15692)"					
BASE COUNT	135 a 229 c 280 g 160 t					
ORIGIN						
alignment_scores:						
Quality:	356.00	Length:	162			
Ratio:	2.781	Gaps:	0			
Percent similarity:	79.012	Percent identity:	41.975			
alignment_block:						
US-09-769-787-162 x U80962/rev						
Align seg 1/1 to reverse of: U80962 from: 1 to: 804						
19	IleGluAsnLeuAsnIleThrIleProIysGlySerLeuPheThrLeuLeu	35				
486	CTTGACGATGTGTCCTGGATATTCACAGCGCAACGAAATCTTCACCTCTCT	437				
35	UGlyAlaSerGlyGlyGlyIstThrIleuLeuArgMetIleAlaGly	52				
436	CGGCGCTTCGGGCTGCGGCAACGACGACGCTCTCCGCTCAACGCGGCT	387				
52	HeAsnSerIleGluGlyGlyGluPheTyrPheAspThrIstLeu	68				
386	TCGACACGCCGATTCGCGACACATCCGCTATAGCGACACACTGCAG	337				
69	AsnMetGluProSerIstLysArgAsnIleGlyMetValPheGlnAsnTyrAl	85				
336	GCGCTGCGCGCTTCGCGCGCGGCTCAACACGAGTTTCCAGAGTTATGC	287				
85	AlIlePheProHisIleuThrValArgAspAsnValAlaPheGlyLeuMetG	102				
286	GCTGTTCGCCGACATGAGCGTGCAGCAACATCGGCTTCGCGCTGCAGGA	237				
102	InLysLysValProIysGluGluIleuIleGlnIstAsnTyrLeu	118				
236	TCGACGCGCGCGCGCGCGACGAGATCGACGCCACCGTGAAGCGCATGCTC	187				
119	GluLeuMetGlnIleAlaGlnTyrAlaAspArgIstProAspLysLeu	135				
186	GACCTGTGGAACCTGCGGAGTCGCGCAAGCGCGCGCGCAACCAATTGTC	137				
135	IGlyGlyGlnGlnArgValThrLeuAlaCysAlaLeuAlaValAsp	152				
136	CGGTGCGCAGCAGCAACGATCGCGCTGCGCGCGCGCTGCGACCGCGC	87				
152	roSerValLeuLeuMetLaspGluProLeuSerAsnLeuGlnAlaLysLeu	168				
86	CCAAAGTGTGCTGCTGACAGATGCGTGTGCGCGCGCTGACCTGAGACTG	37				

```

169 ArgleasmpcArGlnAlaIleagcguilegln 180
169 ::::::::::::::::::::
36 CCCAAGAAATGCAATCGACTGAAGCCCTCGAC 1
36 ::::::::::::::::::::

seq_name: gb_gss:A2049263

seq_documentation_block:
LOCUS      A2049263              405 bp      DNA      linear      GSS 06-MAR-2001
DEFINITION GSSBru0815 Sheared genomic library Brucella melitensis biovar
ABORTUS
ABORTUS genomic clone B16, DNA sequence.
ACCESSION  A2049263
VERSION    A2049263.1  GI:7273178
KEYWORDS
SOURCE     GSS.
ORGANISM   Brucella melitensis biovar Abortus.
            Brucella melitensis biovar Abortus
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Brucellaceae; Brucella.
REFERENCE  1 (bases 1 to 405)
            Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Plerrou,E.,
            Faccoli,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasci,A.C.C.,
            Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.
            Gene discovery through genomic sequencing of Brucella abortus
            Infect. Immun. 69 (2), 865-868 (2001)
TITLE      Contact: Zandomeni, R.
JOURNAL    Centro de Investigacion en Ciencias Agropecuarias (CICA)
            Instituto Nacional de Tecnologia Agropecuaria (INTA) C.C. 25 (1712)
            Castelar, Buenos Aires, Argentina
            Tel: 5411-4621-3316/1683
            Fax: 5411-4481-1316
            Email: zandomeni@inta.gov.ar
COMMENT    Class: shotgun.
            Location/Qualifiers
FEATURES
            1..405
            source
            /organism="Brucella melitensis biovar Abortus"
            /strain="S-2308"
            /db_xref="taxon:235"
            /clone="B16"
            /clone_id="Sheared genomic library"
            /note="Vector: plusscript SK(-) (STRATAGENE); Genomic DNA
            was mechanically sheared, blunt ended, and
            size-fractionated by agarose gel electrophoresis.
            Fragments between 1.5-3 Kb were recovered and ligated to
            the EcoRV site of the plusscript SK (-) vector."
BASE COUNT      79 a      125 c      122 g      77 t      2 others
ORIGIN
alignment_scores:
            Quality: 325.00      Length: 133
            Ratio: 3.037      Gaps: 1
Percent Similarity: 80.451      Percent identity: 48.120
alignment_block:
US-09-769-162 x A2049263 ..
Align seg 1/1 to: A2049263 from: 1 to: 405
79 MetValPheGlnAsnTyraIlePheProHISLeuThrValArgAspAs 95
79 ::::::::::::::::::::
6 ATGCGTTCACATCTATCGCTATCGCATATGAGCGTCCGACAGA 55
6 ::::::::::::::::::::
95 nValAlaPheGlyLeuMetGlnLysLysValPProLysGluGluLeuIleG 112
95 ::::::::::::::::::::
56 CATGGGTTTTCGCTGAARATTCGAACAGCAGCGCGGAGGAAATCGAC 105
56 ::::::::::::::::::::
112 InGlnThrAsnLysTyriLeuGluLeuMetGlnIleAlaGlnTyraLaaSP 128
112 ::::::::::::::::::::
106 GCCGGCGTCAACGAGCGCGCGCTGTTCTGGGGCTTGAGACATGATGAGC 155
106 ::::::::::::::::::::
129 ArgLysProAlaPlyLysLeuSerGlyGlnGlnGlnArgValThrLeuAl 145
129 ::::::::::::::::::::
156 CGCCGCCCGCCGACCTTTCGGGGCGGCGGCGGCGGCGGCGGCGGCGG 205
156 ::::::::::::::::::::

```



COMMENT Contact: McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3099 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@fscs.sdu.edu  
Class: shotgun

FEATURES  
source  
1.1340  
/organism="Salmonella typhimurium"  
/strain="LT2"  
/db\_xref="taxon:602"  
/clone="593-F7"  
/clone\_1lb="Salmonella typhimurium LT2, Lambda DASH II"  
/note="Vector: Lambda DASH II; sequenced using LI-Cor  
sequencer"

BASE COUNT 308 a 382 c 293 g 291 t 66 others

ORIGIN

alignment\_scores:  
Quality: 320.50 Length: 227  
Ratio: 2.016 Gaps: 6  
Percent Similarity: 70.044 Percent Identity: 36.123

alignment\_block:  
US-09-769-787-162 x AF029490/rev ..

Align seg 1/1 to reverse of: AF029490 from: 1 to: 1340

```

70 MetCupProserLysArgAsnIleGlyMetValPheGlnAsnTyrrAlaIle 86
1056 ATGTCGCCCGCATGACGGCTGATTAATGATGTTTCAGTCNTATGCGCTG 1007
86 ephProHisLeuThrValArgAspAsnValAlaPheGlyLeuMetGlnL 103
1006 T...TCCGATATGACGTA.....GACGACATCGCTTGTCTCGA 970
103 yslLysValProLysGluGluLeuIleGlnGlnThrAsnLysTyrLeuGlu 119
969 GNAAGANANTGCGAAGCGAG...ATACAGCGCGTCAATGAGATGCTGAT 923
120 LeuMetGlnIleAlaGlnTyrrAlaAspArgLysProAspLysLeuSerG 136
922 CTGGTGCACATG...CAGAGTTGCGAAACGTAGCGCGCATGTHNCCGG 876
136 yGlyGlnGlnIleArgValThrLeuAlaCysAlaLeuAlaValAsnProS 153
875 CGNTCAGCGTACGGCGCTGGCCCTGGCGCGCAGCTGGCGGAAGCTCTTA 826
153 eValLeuLeuMetAspGluProLeuSerAsnLeuGluAlaLysLeuArg 169
825 AGCGCTCTTGTGTGATGACCGATGGCGCGCGCATTAATAAATTTACGC 776
170 LeuAspMetArgGlnAlaIleArgGlnIleGlnIleGlnIleGlyIleTh 186
775 GACGAGATGACGCTGGAAGTGTGATATCTTWAAGCGCGCTGAC 726
186 rThrValTyrValThrHisAspGlnGluGluAlaMetAlaIleSerAspG 203
725 CTGGCTATGATGATACGACGATCAGAAAGCATGATGATGCGGGGC 676
203 InitAlaValMetLysAspGlyValIleGlnGlnIleGlyArgProLys 219
675 GCATTTGCCATCATGACCGTGAATAATTTGTCAAAATTTGCGAAGCGAG 626
220 GluLeuTyrHisLysProAlaAsnGluPheValAlaThrPheIleGlyAr 236
625 GAGATTATACGACATCCGACACCCGCTACAGCGCGCAATTTATCGGATC 576
236 gThrAsnIleIleProAlaAsnLeuGluLysArgSerAspGlyValAlaTyr 253
575 GGTAAATGTTTTCAGGCTTACTGAAGCCGCTGAGAGAGAGCGTCTGC 526

```

253 leValPheSerAspGly.....TyrAlaLeuArgMetProAla...Leu 266  
525 TCATTGACGCCCGCGGTTTGGTTCATCGTTAAAGTCGATGACGATGCC 476  
267 AspGlnValGluGlnGlnAlaIleHisValSerIleArgProGluGluPh 283  
475 TCTGCTGTTGATTAATGTCGCCGCTATNTGTGGCTACGTCGGAANAAT 426  
283 eileLysAspGluSerGlyAspIleGlyGly 293  
425 CATGCTGTGATGATGTCGCCGCTCGCATGCG 395  
seq\_name: gb\_est1:A1058144

seq\_documentation\_block:  
LOCUS A1058144 417 bp mRNA linear EST 20-JUL-1998  
DEFINITION SMOVL3CAN22H02SK Onchocerca volvulus infective larva cDNA  
(SAM94WL-OvL3) Onchocerca volvulus cDNA clone SMOVL3CAN22H02 5',  
mRNA sequence.

ACCESSION A1058144  
VERSION A1058144.1 GI:3332010  
KEYWORDS EST.  
SOURCE Onchocerca volvulus.  
ORGANISM Onchocerca volvulus.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
Onchocercidae; Onchocerca.  
1 (bases 1 to 417)  
Williams, S.A., Lu, W., Lizotte-Waniewski, M. and Laney, S.J.  
Genes expressed in infective third stage larvae of Onchocerca  
volvulus  
Unpublished (1995)  
Contact: Steven A. Williams  
Molecular Parasitology  
Smith College Department of Biological Sciences  
Department of Biological Sciences, Clark Science Center, Smith  
College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: genomes@smith.edu  
Seq primer: pbLuescript SK.

FEATURES  
source  
1.417  
Location/Qualifiers  
/organism="Onchocerca volvulus"  
/strain="Stieria Leone"  
/db\_xref="taxon:6282"  
/clone="SMOVL3CAN22H02"  
/clone\_1lb="Onchocerca volvulus infective larva cDNA  
(SAM94WL-OvL3)"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda UniZap XR, Site\_1: EcoR I; Site\_2:  
Xho I; Cutaneous filarial nematode parasite of humans.  
mRNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNAPol I. The library had 1.8 x 10<sup>5</sup> independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genomes@smith.edu."

BASE COUNT 93 a 115 c 108 g 94 t 7 others

ORIGIN

alignment\_scores:  
Quality: 315.00 Length: 126  
Ratio: 3.029 Gaps: 1  
Percent Similarity: 82.540 Percent Identity: 50.000

alignment\_block:  
US-09-769-787-162 x A1058144 ..

Align seg 1/1 to: A1058144 from: 1 to: 417



alignment\_scores:

1  
2  
3

4  
5  
6

---